Docket: 33390/US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	Yonglian Zhang		
Appln. No.:	10/623,629		
Filed:	July 21, 2003	Examiner:	R. Mondesi
Title:	A Novel Natural Antibacterial Peptide, the Nucleotide Sequence Encoding it and the Use Thereof	Group Art Unit:	1653

LETTER SUBMITTING CERTIFIED COPY PURSUANT TO 35 U.S.C. §119

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

I hereby certify that this document is being sent via First Class U. S. mail addressed to Commissioner for Patents, P. O. Box 1450, Alexandria, VA 22313-1450, on this 28 day of October, 2004.

Frances Cook
(Signature)

Dear Sir:

Pursuant to 35 U.S.C. §119, to perfect the claim for foreign priority benefits in the above-identified patent application, enclosed for filing is a certified copy of Chinese Application No. 01005283.X, as filed on January 22, 2001, including specification ,drawings, and Verification of Translation.

Respectfully submitted,

DORSEY & WHITNEY LLP
Customer Number 25763

Date: October 28, 2004

By: S. D. Solberg

Sean D. Solberg, (Reg. No. 48,653)
Intellectual Property Department
Suite 1500, 50 South Sixth Street
Minneapolis, MN 55402-1498
(612) 340-7862

Priority document

This is to certified that the attached document is a true copy of the application document filed with the Chinese Patent Office.

Filing date:	Jan. 22, 2001
Application No.:	01105283.X
Application type:	Utility invention
Title of invention:	A NOVEL NATURAL ANTIBACTERIAL PEPTIDE, THE NUCLEOTIDE SEQUENCE ENCODING IT AND THE USE THEREOF
Applicant(s):	SHANGHAI INSTITUTION OF BIOCHEMISTRY, CAS
Inventor(s):	(1) ZHANG, Yonglian (2) LI, Peng (3) HE, Bin (4)SHANG, Quan (5)ZHANG, Youduan (6) CHAN, Hsiaochang (7)SO, Siucheung (8)CHUNG, Yiuwa

Chinese patent Office
Director : WANG, Jinchuan
September 23, 2004

证　　明

本证明之附件是向本局提交的下列专利申请副本

申　　日：　2001.01.22

申　　号：　01105283.X

申 请 类 别：　发明

发明创造名称：　新的天然抗菌肽、其编码序列及用途

申　请　人：　中国科学院上海生物化学研究所

发明人或设计人：　张永莲、陈小章、李鹏、何彬、苏兆祥、钟耀华、商权、张友端

CERTIFIED COPY OF
PRIORITY DOCUMENT

中华人民共和国
国家知识产权局局长　王景川

2004 年 9 月 23 日

权 利 要 求 书

1. 一种分离或纯化的多肽，其特征在于，它包括具有SEQ ID NO: 2或3所示氨基酸序列的多肽。

2. 如权利要求1所述的多肽，其特征在于，该多肽具有SEQ ID NO: 2或3所示氨基酸序列。

3. 一种分离或纯化的多核苷酸，其特征在于，它包含一核苷酸序列，该核苷酸序列与选自下组的一种核苷酸序列有至少70%相同性：

(a)编码SEQ ID NO:2或3所述多肽的多核苷酸；

(b)与多核苷酸(a)互补的多核苷酸。

4. 如权利要求3所述的多核苷酸，其特征在于，该多核苷酸编码具有SEQ ID NO: 2或3所示氨基酸序列的多肽。

5. 一种载体，其特征在于，它含有权利要求3所述的多核苷酸。

6. 一种遗传工程化的宿主细胞，其特征在于，它含有权利要求5所述的载体。

7. 一种具有Binlb蛋白活性的多肽的制备方法，其特征在于，该方法包含：

(a)在适合表达Binlb蛋白的条件下，培养权利要求6所述的宿主细胞；

(b)从培养物中分离出具有Binlb蛋白活性的多肽。

8. 一种能与权利要求1所述的Binlb蛋白特异性结合的抗体。

9. 一种药物组合物，其特征在于，它含有安全有效量的权利要求 1 所述的多肽以及药学上可接受的载体。

10. 一种杀菌剂，其特征在于，它含有有效量的权利要求 1 所述的多肽。

01.02.05

6

说　明　书

新的天然抗菌肽、其编码序列及用途

本发明涉及分子生物学、生殖生物学、分子免疫学和医学领域。具体地说，本发明涉及新的在大鼠附睾头部特异表达的天然抗菌肽-Bin1b 蛋白，以及编码该抗菌肽的多核苷酸。本发明还涉及此抗菌肽及其编码序列的制备方法和用途。Bin1b 蛋白还与男性生育，尤其是精子成熟有关。

人类基因组全顺序的阐明会绘出一篇密码图谱，随后的是了解哪一段顺序编码哪一个基因?这一基因在体内的功能是什么?一个生物学功能需要哪几个基因的协同作用?它们又是如何协调控制着生命的正常活动的?以及哪些错误会导致各种病理现象的产生?这一系列艰巨的，极富挑战性的，能直接为人类造福的，也能产生巨大经济效益的解码工作促使人们进入了功能基因组研究时代。

在回顾 20 世纪和展望 21 世纪之际，国际同行专家的一致认为：生殖生物学发展至今虽然成果颇丰，但如果沿袭过去的思路和方法，局限于组织水平、细胞水平的观察和测定，不借鉴相关学科的新技术，不深入研究生殖的分子机制，不进行多学科的交叉研究，那么其发展将不容乐观。而且，多年来生殖生物学主要侧重于雌性生殖的研究，对雄性生殖领域的涉足不多。人们对揭示生殖规律的基础研究尚不够重视，致使迄今尚无理想的节育药物和技术，且对自身生殖奥秘的认识滞后于现代生命科学的发展。因而生殖生物学在 21 世纪将面临两项革命：(1)重视发展其基础研究，将现代分子生物学、分子免疫学和细胞生物学等新思想、新方法、深入研究控制人类生殖和出生缺陷的机制；(2)加强雄性生殖生物学的研究，大力发展男性避孕措施，为当前生殖健康研究的国际趋势。

中国人口目前占世界人口的 22%，但耕地只有世界的 7%，淡水量只有世界的 6%。即便将人口增长率控制在 0.01%，每年也将有 0.13 亿以上的人口逐年增长。此外，目前所使用的避孕措施主要以女性为主。何况这些避孕药还未完善，有着这样或那样的副作用(其遗传后果还犹待评估)。安全的绝育措施也以女性为主，据 WHO 的统计，我国四川省男性输精管结扎者只有 1/12。实际上，避孕是男女双方都有责任的，而且男性生殖调控在人口数量和质量控制两个方面都更具重要性。因为(1)健康男性在 50 年生育期内，每天产生 1 亿精子，而女性生育期只有 40 年，每个月才排出 1 个成熟卵。(2)精子容易受环境影响而产生突变。(3)50 年来，精子数量和质量下降了 40%。(4)45 岁以下的男性原发性不育者占 5-10%。因此，除了要纠正“避孕是妇女的事”这一误区外，还要特别加强雄性生殖生物学的研究，大力发展男性避孕措施。众所周知，人类生殖健康是全球性战略的头

等大事。避孕毕竟是缓解人口危机的权宜之计，而后代在本星球上追求身心健康的生活方式才是恒久主题，生育调节药物的设计应适应新的现实需求，更上一层楼。1999 年 9 月 9-10 日美国儿童健康与人类发育研究所(NICHD)负责在 NIH 总部召开了“男性避孕在 21 世纪”研讨会。交流了现状，提出了目标，布署了措施(组织合作及基金支持等)，打响了战鼓[Trends in Endocrinology and Metabolism 2000, 11(2):66-69].

目前已知精祖细胞在睾丸中通过有丝分裂、减数分裂和形态学分化成为完全的精子细胞后，进入附睾，在其头部和体部逐渐达到成熟，然后在尾部储存，等待射精。而精子运动能力的形成、顶体功能的完善、代谢类型的转变等一系列程序化的成熟功能的获得并非靠其自身完成，而是在其通过附睾的过程中，与附睾微环境相互作用中按其特定的程序逐渐成熟的。附睾是一根连接睾丸和输精管的盘曲的细长管道。但其每一段落的附睾细胞表达不同的基因及基因产物，分泌种类不同的蛋白及免疫抑制分子、有不同的腔液组成、一定的离子强度、pH 等，形成不断变化着的管腔微环境，与精子相互作用，使精子表面蛋白的成分得以部分改变或得以部分修饰(如加减一序列的磷酸化，酯化，酰化，羧基化，醣基化等)而被激活或抑制从而逐步获得成熟时所具备的功能及免疫防御能力，保护其自身能在附睾中储存及通过女性生殖道直至与卵结合。

附睾的这种对精子成熟、储存和保护的作用因具有下述诸特点而早已被认同并视作进行抗生育的一个理想的靶器官：

(1)功能较为单一。干扰其功能的药物不致引起严重的副反应。

(2)激素作用的终极器官。自身不具有内分泌机能。故作用于附睾的药物一般不会影响激素分泌。

(3)精子在进入附睾前已分化完全，基因转录已停止，其成熟功能的获得一般与蛋白的修饰有关，不涉及 DNA 复制机能。因此药物的作用不可能引起 DNA 改变的遗传病。

(4)附睾的研究尚未引起人们的足够重视，因此是一个未被开垦的处女地，大有用武之地。

然而，附睾不同部位形成这些程序化微环境差异的机制绝不是一、两个基因和蛋白所能左右的，而是一组组基因程序性表达产物的协同作用的结果。而对于这种与精子成熟相关的附睾特异基因表达的启动与程序还知之甚少。因此对这方面的研究，不仅有助于揭示精子在附睾中成熟的分子机理，解读一部分基因组密码，而且还为解决精子成熟异常所引起的不孕症提供分子基础，也为开发一些阻断精子成熟的男性避孕药物提供新的设计思路。

纵观文献，对附睾的研究早已受到关注。回顾自 70 年代至今，主要从三个方面在进行探索：

(1)从蛋白质水平而言，利用双向蛋白质电泳等技术将附睾不同部位的管腔中

的蛋白质或不同部位成熟程度不同的精子上的膜蛋白进行比较，或用这些不同蛋白质制成多抗，用免疫法来进行比较取得了些结果，然而由于受分离分析技术的灵敏度的限制，进展不尽如人意。近年来，由于蛋白组学研究的突飞猛进，法国 Dacheux 实验室自 96 年到今年初报道在猪和羊附睾的管腔中鉴定到 200 多个蛋白。提示与其它器官相比，附睾研究的难度相对较小。然而迄今为止，只有 15 个附睾特异的蛋白的 cDNA 被克隆。

(2)通过已知附睾在精子成熟过程中的作用，来研究一些已知功能的蛋白质，是否与附睾某一特定功能相关。如已知附睾能保护精子免受氧自由基的损伤。曾对附睾不同部位 6 种抗氧化酶 mRNA 进行检测，发现在其头部 E-GPX 的 mRNA 含量最高，而在体部则 E-SOD 的 mRNA 含量最高。这些结果提示在附睾不同部位所需的抗氧化酶的种类不一。然而这方面的研究局限在已有的知识基础上，对新基因产物或新功能基因的发现上则很少有帮助。另外，已知附睾内存在大量免疫细胞与免疫抑制分子，保护精子不受自身攻击的免疫微环境，但其形成与筛选调控的分子和细胞机制不清，资料不系统。

(3)90 年代以来由于分子生物学的技术日新月异，开始有可能利用差减杂交等方法从 mRNA 水平上来搜寻附睾特异表达的新基因。如西德实验室用差示筛选 cDNA 库的方法在人附睾中先后找到 6 个新 mRNA，然而由于人类附睾或其它的材料局限，无法对其功能进行深入研究，因而都在用不同手段在其它实验动物中进行搜索。

因此，本领域迫切需要开发新的与男性生育有关和/或与附睾有关的天然蛋白。

本发明的目的是提供一种新的天然抗菌肽-Binlb 蛋白以及其片段、类似物和衍生物。

本发明的另一目的是提供编码这些多肽的多核苷酸。

本发明的另一目的是提供生产这些多肽和多核苷酸的方法及其用途。

在本发明的第一方面，提供新颖的分离或纯化的 Binlb 多肽，该多肽源自大鼠，它包含：具有 SEQ ID NO: 2 或 3 氨基酸序列的多肽、或其保守性变异多肽、或其活性片段、或其活性衍生物。较佳地，该多肽是具有 SEQ ID NO: 2 或 3 氨基酸序列的多肽。

在本发明的第二方面，提供了编码分离或纯化的这些多肽的多核苷酸，该多核苷酸包含一核苷酸序列，该核苷酸序列与选自下组的一种核苷酸序列有至少 70% 相同性：(a)编码 SEQ ID NO: 2 或 3 所示 Binlb 多肽的多核苷酸；和(b)与多核苷

酸(a)互补的多核苷酸。较佳地，该多核苷酸编码具有 SEQ ID NO: 2 或 3 所示氨基酸序列的多肽。更佳地，该多核苷酸的序列是选自下组的一种：(a)具有 SEQ ID NO: 1 中 57-263 位的序列；(b)具有 SEQ ID NO: 1 中 105-263 位的序列；(c)具有 SEQ ID NO: 1 中 1-336 位的序列。

在本发明的第三方面，提供了含有上述多核苷酸的载体，以及被该载体转化或转导的宿主细胞或者被上述多核苷酸直接转化或转导的宿主细胞。

在本发明的第四方面，提供了制备具有 Binlb 蛋白活性的多肽的方法，该方法包含：(a)在适合表达 Binlb 蛋白的条件下，培养上述被转化或转导的宿主细胞；(b)从培养物中分离出具有 Binlb 蛋白活性的多肽。

在本发明的第五方面，提供了与上述的 Binlb 多肽特异性结合的抗体。还提供了可用于检测的核酸分子，它含有上述的多核苷酸中连续的 15-1757 个核苷酸。

在本发明的第六方面，提供了模拟、促进、拮抗 Binlb 多肽活性的化合物，以及抑制 Binlb 多肽的表达的化合物。还提供了筛选和/或制备这些化合物的方法。较佳地，该化合物是 Binlb 多肽的编码序列或其片段的反义序列。

在本发明的第七方面，提供了检测样品中是否存在 Binlb 蛋白的方法，它包括：将样品与 Binlb 蛋白的特异性抗体接触，观察是否形成抗体复合物，形成了抗体复合物就表示样品中存在 Binlb 蛋白。

在本发明的第八方面，提供了一种检测与 Binlb 多肽异常表达相关的疾病或疾病易感性的方法，该方法包括：检测编码所述多肽的核酸序列中是否存在突变。

在本发明的第九方面，提供了本发明多肽和编码序列的用途。例如本发明多肽可被用于筛选促进 Binlb 多肽活性的激动剂，或者筛选抑制 Binlb 多肽活性的拮抗剂、或者被用于肽指纹图谱鉴定。本发明的 Binlb 蛋白的编码序列或其片段，可被作为引物用于 PCR 扩增反应，或者作为探针用于杂交反应，或者用于制造基因芯片或微阵列。

在本发明的第十方面，提供了一种药物组合物，它含有安全有效量的本发明的 Binlb 多肽或其激动剂、拮抗剂以及药学上可接受的载体。这些药物组合物可治疗泌尿生殖系统感染等病症。

在本发明的第十一方面，提供了一种杀菌剂，它含有杀菌有效量的本发明 Binlb 多肽。

本发明的其它方面由于本文的技术的公开，对本领域的技术人员而言是显而易见的。

本发明人利用 mRNA 的差异显示分析及差减杂交库分析分别对大鼠和猴附睾不同部位特异表达的基因作了筛选(猴方面的工作与美国北卡大学合作)。不仅鉴定

到了那些已被发现的特异基因，而且还获得一系列的特异表达的新基因的全长cDNA克隆(2个在大鼠附睾头部、4个在猴附睾头部、4个在猴附睾体部、3个在猴附睾尾部)。

Binlb是其中之一，它是在大鼠附睾头部特异表达的基因，已取得其全长cDNA和基因组DNA的克隆(其核苷酸和氨基酸序列和结构已在美国NIH基因银行中登陆，登陆号为AF217088， AF217089。在本申请之前，这些序列还未公开)。Binlb基因的表达非常特异，只在大鼠附睾头部的上皮细胞中产生，在大鼠生育旺盛期表达最高，年老后下降。提示可能与生育有关。而且，Binlb基因的表达受雄激素正调控。这为今后利用激素小分子来影响此基因表达以调控精子成熟设计男性避孕药物成为可能。此外，Binlb基因具有抗菌作用，是在大鼠附睾中发现的第一个beta-防卫素类天然抗菌肽，有望发展成为治疗泌尿生殖系统感染的一个天然药物。

下列附图用于说明本发明的具体实施方案，而不用于限定由权利要求书所界定的本发明范围。

图1显示了大鼠附睾头(1)、体(2)和尾(3)的DD-RT-PCR电泳图。样品一式两份重复以确保准确性。如箭头所示，Binl在头部区域是差异表达的。

图2显示了Binlb的基因组序列和结构特性。Binlb的基因组DNA序列(Genebank登录号AF217089)，是通过使用位于Binlb全长cDNA序列(AF217088)两端的引物(上游引物：5'- GGACACCCAGTCATCAGTCACAT-3'(SEQ ID NO: 9)和下游引物：5'- TTTGGGGTGCTT CCAGGTCTCT-3'(SEQ ID NO: 10))，并以大鼠基因组DNA为模板PCR扩增而得的。两个外显子：用大写字母表示；编码区：用阴影表示；poly A结尾信号：粗体大写字母；内含子：小写字母；剪接位点：粗体小写字母；推定的信号肽：带下划线的氨基酸。在N端有一豆蔻酰化的G残基(用方框表示)，其共有模式(GIRNTV)用粗斜体大写字母标出。圆圈中的S残基可以被PKC磷酸化，其共有模式(SIK)用粗斜体大写字母标出。终止密码子用“*”表示。

图2B显示了Binlb与β-防卫素的序列相似性。保守的6个浅灰阴影标出。BNBD9(AAB25872)， BNBD3(AAB25866)， BNBD7(AAB25870)， TAP(P25068)，LAP(Q28880)， EBD(002775) 来自牛； HBD1(Q09753)， HBD2(015263)，HBD3(NP061131) 来自人； EP2E(AF263555_1) CBD1(AF188607_1)，CBD2(AF209855_1) 来自黑猩猩； MBD1(AAB72003)， MBD2(CAB42815)，MBD3(AF092929_1)， MBD4(AF155882_1) 来自小鼠； RBD1(AAC28071)，RBD2(AAC28072) 来自大鼠； GBD1(CAA76811)， GBD2(CAA08905) 来自山羊；SBD1(019038)，SBD2(019039) 来自绵羊；PBD1(062697) 来自猪；Gal1(P46156)，Gal1a(P46157)和Gal2(P46158) 分别是鸡的gallinacin 1, 1α和2；THP1(P80391)

和 THP2(P80392)分别是火鸡的异嗜性肽 1 和 2. EP2E 是黑猩猩中与 Binlb 同源的一种同源蛋白，然而其作者认为它不是β-防卫素家族的成员。

图 2C 是 Binlb 与灵长目同源蛋白的序列比较。与 Binlb 相比，灵长目同源蛋白 EP2D(AF263554_1)和 HE2β1(AF168617_1)具有更长的 N 端和 C 端。EP2E(AF262555_1) 仅具有更长的 C 端。在图中，HE2β1 用 HE1b1 表示。

图 3 显示了 Binlb 全长 cDNA 的体外转录和翻译试验的结果。

图 3A 是对 Binlb 全长 cDNA 的体外转录和翻译产物的分析电泳图。泳道 1 和 2 为一式两份的重复样品。

图 3B 是用与图 3A 中相同的系统，对阴性对照(无质粒 DNA)和荧光素酶 SP6 和 T7 对照进行分析的电泳图。其中，因分子量差异而在 12%SDS-PAGE 中电泳。泳道 1 为阴性对照(无质粒 DNA)，泳道 2 为荧光素酶 SP6 对照，泳道 3 为荧光素酶 T7 对照。

图 3C. 质粒 pSPT18-Binlb 的结构图。

图 4 显示了重组的 Binlb 融合蛋白的表达。融合蛋白 DHFR-Binlb 的分子量为 31KD，如箭头所指。各泳道如下：NI:未诱导，I:诱导，CL: 澄清裂解液，FT: 流穿液，W:洗涤缓冲液，E1:洗脱 1，E2:洗脱 2，M: 分子量标准。

图 5 显示了 Binlb 的定位以及发育调控。(A). 用 Northern 分析确定的 Binlb 组织分布情况。(B). 用原位杂交确定的 Binlb 在附睾中的区域分布情况，Binlb 位于大鼠附睾头部区域的中间(Mid)。Ins: 初始段。(C)用反义探针确定的 Binlb 细胞定位。Binlb 位于附睾上皮细胞(Prc)中。(D). 有义探针。(E)作为阳性对照的 18s 探针。(E). 在整个生命期间(15-720 天)，大鼠附睾中 Binlb mRNA 的表达图。(泳道 1，2 和 3 分别表示来自附睾头、体、尾区域的 mRNA)。

图 6. 在用 EDS(Ethylene Dimethamesulfonate)处理后，对大鼠头部总 RNA 进行的 Northern 印迹分析结果。

图 6A 为 Northern 印迹电泳杂交图。

图 6B 显示了 Binlb 的表达受雄激素的上调。其中，带“▲”的曲线表示在注射 EDS 后间隔不同时间，大鼠血清中的睾丸激素的水平；带“◆”的曲线表示了用 18s 水平校正后大鼠 Binlb 表达水平的变化。

图 7 显示了 Binlb 的抗菌活性以及对炎症作出的表达上调。(A). 头部和尾部培养物抗菌活性的比较。在试验前 16 小时，将 100 CFU 大肠杆菌加至培养物中。(B). 反义 Binlb 对头部培养物抗菌活性的影响。用反义或有义寡核苷酸(5 微克/微升)转染培养物 20 小时。(C). 通过结扎大鼠输精管而引起炎症(2 周)，在附睾头部 Binlb mRNA 增加，而尾部没有变化。

在本发明中，术语“Binlb 蛋白”、“Binlb 多肽”或“抗菌肽-Binlb”可互

换使用，都指具有天然抗菌肽 Binlb 氨基酸序列(SEQ ID NO:2 或 3)的蛋白或多肽。它们包括含有或不含起始甲硫氨酸的天然抗菌肽 Binlb，以及含有或不含有信号肽的 Binlb 蛋白。成熟的 Binlb 蛋白的氨基酸序列示于 SEQ ID NO: 3。

如本文所用，“分离的”是指物质从其原始环境中分离出来(如果是天然的物质，原始环境即是天然环境)。例如，活体细胞内的天然状态下的多聚核苷酸和多肽是没有分离纯化的，但同样的多聚核苷酸和多肽如果与天然状态一起存在的其他物质分开，则为分离纯化的。“分离”、“纯化”包括将重组表达的 Binlb 蛋白与其他蛋白、糖类等物质分开。

如本文所用，“分离或纯化的 Binlb 蛋白或多肽”是指 Binlb 多肽基本上不含天然与其相关的其它蛋白、脂类、糖类或其它物质。本领域的技术人员能用标准的蛋白质纯化技术纯化 Binlb 蛋白。

本发明的多肽可以是重组多肽、天然多肽、合成多肽，优选重组多肽。本发明的多肽可以是天然纯化的产物，或是化学合成的产物，或使用重组技术从原核或真核宿主(例如，细菌、酵母、高等植物、昆虫和哺乳动物细胞)中产生。根据重组生产方案所用的宿主，本发明的多肽可以是糖基化的，或可以是非糖基化的。本发明的多肽还可包括或不包括起始的甲硫氨酸残基。

本发明还包括 Binlb 蛋白的片段、衍生物和类似物。如本文所用，术语“片段”、“衍生物”和“类似物”是指基本上保持本发明的天然 Binlb 蛋白相同的生物学功能或活性的多肽。本发明的多肽片段、衍生物或类似物可以是(i)有一个或多个保守或非保守性氨基酸残基(优选保守性氨基酸残基)被取代的多肽，或(ii)在一个或多个氨基酸残基中具有取代基团的多肽，或(iii)成熟多肽与另一个化合物(比如延长多肽半衰期的化合物，例如聚乙二醇)融合所形成的多肽，或(iv)附加的氨基酸序列融合于此多肽序列而形成的多肽(如前导序列或分泌序列或用来纯化此多肽的序列或蛋白原序列，或与抗原 IgG 片段的形成的融合蛋白)。根据本文的教导，这些片段、衍生物和类似物属于本领域熟练技术人员公知的范围。

一类特殊的 Binlb 类似物是在其他哺乳动物(如牛、羊、兔、狗、猴、人等)中 Binlb 的同源蛋白。这些其他物种的同源蛋白的编码序列，可根据本发明公开的序列，通过杂交或扩增的方法而获得，进而通过常规重组方法获得这些同源蛋白。

在本发明中，术语“Binlb 多肽”指具有 Binlb 蛋白活性的 SEQ ID NO. 2 或 3 序列的多肽。该术语还包括具有与 Binlb 蛋白相同功能的、SEQ ID NO. 2 或 3 序列的变异形式。这些变异形式包括(但并不限于)：若干个(通常为 1-20 个，较佳地 1-10 个，更佳地 1-5 个，最佳地 1-3 个)氨基酸的缺失、插入和/或取代，以及在 C 末端和/或 N 末端添加一个或数个(通常为 20 个以内，较佳地为 10 个以内，更佳地为 5 个以内)氨基酸。例如，在本领域中，用性能相近或相似的氨基酸进行取代时，

通常不会改变蛋白质的功能。又比如，在C末端和/或N末端添加一个或数个氨基酸通常也不会改变蛋白质的功能。该术语还包括Binlb蛋白的活性片段和活性衍生物。

该多肽的变异形式包括：同源序列、保守性变异体、等位变异体、天然突变体、诱导突变体、在高或低的严紧度条件下能与Binlb DNA 杂交的DNA所编码的蛋白、以及利用抗Binlb多肽的抗血清获得的多肽或蛋白。本发明还提供了其他多肽，如包含Binlb多肽或其片段的融合蛋白。除了几乎全长的多肽外，本发明还包括了Binlb多肽的可溶性片段。通常，该片段具有Binlb多肽序列的至少约15个连续氨基酸，通常至少约25个连续氨基酸，更佳地至少约35个连续氨基酸，最佳地至少约40个连续氨基酸。

发明还提供Binlb蛋白或多肽的类似物。这些类似物与天然Binlb多肽的差别可以是氨基酸序列上的差异，也可以是不影响序列的修饰形式上的差异，或者兼而有之。这些多肽包括天然或诱导的遗传变异体。诱导变异体可以通过各种技术得到，如通过辐射或暴露于诱变剂而产生随机诱变，还可通过定点诱变法或其他已知分子生物学的技术。类似物还包括具有不同于天然L-氨基酸的残基(如D-氨基酸)的类似物，以及具有非天然存在的或合成的氨基酸(如β、γ-氨基酸)的类似物。应理解，本发明的多肽并不限于上述例举的代表性的多肽。

修饰(通常不改变一级结构)形式包括：体内或体外的多肽的化学衍生形式如乙酰化或羧基化。修饰还包括糖基化，如那些在多肽的合成和加工中或进一步加工步骤中进行糖基化修饰而产生的多肽。这种修饰可以通过将多肽暴露于进行糖基化的酶(如哺乳动物的糖基化酶或去糖基化酶)而完成。修饰形式还包括具有磷酸化氨基酸残基(如磷酸酪氨酸，磷酸丝氨酸，磷酸苏氨酸)的序列。还包括被修饰从而提高了其抗蛋白水解性能或优化了溶解性能的多肽。

在本发明中，“Binlb蛋白保守性变异多肽”指与SEQ ID NO: 2或3的氨基酸序列相比，有至多10个，较佳地至多8个，更佳地至多5个，最佳地至多3个氨基酸被性质相似或相近的氨基酸所替换而形成多肽。这些保守性变异多肽最好根据表1进行氨基酸替换而产生。

表 1

最初的残基	代表性的取代	优选的取代
Ala (A)	Val; Leu; Ile	Val
Arg (R)	Lys; Gln; Asn	Lys
Asn (N)	Gln; His; Lys; Arg	Gln
Asp (D)	Glu	Glu
Cys (C)	Ser	Ser
Gln (Q)	Asn	Asn
Glu (E)	Asp	Asp
Gly (G)	Pro; Ala	Ala
His (H)	Asn; Gln; Lys; Arg	Arg
Ile (I)	Leu; Val; Met; Ala; Phe	Leu
Leu (L)	Ile; Val; Met; Ala; Phe	Ile
Lys (K)	Arg; Gln; Asn	Arg
Met (M)	Leu; Phe; Ile	Leu
Phe (F)	Leu; Val; Ile; Ala; Tyr	Leu
Pro (P)	Ala	Ala
Ser (S)	Thr	Thr
Thr (T)	Ser	Ser
Trp (W)	Tyr; Phe	Tyr
Tyr (Y)	Trp; Phe; Thr; Ser	Phe
Val (V)	Ile; Leu; Met; Phe; Ala	Leu

本发明的多核苷酸可以是 DNA 形式或 RNA 形式。DNA 形式包括 cDNA、基因组 DNA 或人工合成的 DNA。DNA 可以是单链的或是双链的。DNA 可以是编码链或非编码链。编码成熟多肽的编码区序列可以与 SEQ ID NO:1 所示的编码区序列相同或者是简并的变异体。如本文所用，“简并的变异体”在本发明中是指编码具有 SEQ ID NO:2 或 3 的蛋白质，但与 SEQ ID NO:1 中相应编码区序列有差别的核酸序列。

编码 Binlb 成熟多肽的多核苷酸包括：只编码成熟多肽的编码序列；成熟多肽的编码序列+各种附加编码序列；成熟多肽的编码序列+任选的附加编码序列+非编码序列。

术语“编码多肽的多核苷酸”可以是包括编码此多肽的多核苷酸，也可以是还包括附加编码和/或非编码序列的多核苷酸。

本发明还涉及上述多核苷酸的变异体，其编码与本发明有相同的氨基酸序列的多肽或多肽的片段、类似物和衍生物。此多核苷酸的变异体可以是天然发生的等位变异体或非天然发生的变异体。这些核苷酸变异体包括取代变异体、缺失变异体和插入变异体。如本领域所知，等位变异体是一种多核苷酸的替换形式，它可能是一个或多个核苷酸的取代、缺失或插入，但这种变化不会从实质上改变其编码的多肽的功能。

本发明还涉及与上述的序列杂交且两个序列之间具有至少 50%，较佳地至少 70%，更佳地至少 80%相同性的多核苷酸。本发明特别涉及在严格条件下与本发明所述多核苷酸可杂交的多核苷酸。在本发明中，“严格条件”是指：(1)在较低离子强度和较高温度下的杂交和洗脱，如 0.2×SSC，0.1%SDS，60℃；或(2)杂交时加有变性剂，如 50%(v/v)甲酰胺，0.1%小牛血清/0.1% Ficoll，42℃等；或(3)仅在两条序列之间的相同性至少在 90%以上，更好是 95%以上时才发生杂交。并且，可杂交的多核苷酸编码的多肽与 SEQ ID NO:2 所示的成熟多肽有相同的生物学功能和活性。

本发明还涉及与上述的序列杂交的核酸片段。如本文所用，“核酸片段”的长度至少含 15 个核苷酸，较好是至少 30 个核苷酸，更好是至少 50 个核苷酸，最好是至少 100 个核苷酸以上。核酸片段可用于核酸的扩增技术(如 PCR)以确定和/或分离编码 Binlb 核苷酸序列。

本发明的Binlb核苷酸全长序列或其片段通常可以用PCR扩增法、重组法或人工合成的方法获得。对于PCR扩增法，可根据本发明所公开的有关核苷酸序列，尤其是开放阅读框序列来设计引物，并用市售的cDNA库或按本领域技术人员已知的常规方法所制备的cDNA库作为模板，扩增而得有关序列。当序列较长时，常常需要进行两次或多次PCR扩增，然后再将各次扩增出的片段按正确次序拼接在一起。

一旦获得了有关的序列，就可以用重组法来大批量地获得有关序列。这通常是将其克隆入载体，再转入细胞，然后通过常规方法从增殖后的宿主细胞中分离得到有关序列。

此外，还可用人工合成的方法来合成有关序列，因为Binlb编码序列的长度较短。通常，通过先合成多个小片段，然后再进行连接可获得序列很长的片段。

目前，已经可以完全通过化学合成来得到编码本发明蛋白(或其片段，或其衍生物)的 DNA 序列。然后可将该 DNA 序列引入本领域中已知的各种现有的 DNA 分子(或如载体)和细胞中。此外，还可通过化学合成将突变引入本发明蛋白序列中。

应用 PCR 技术扩增 DNA/RNA 的方法(Saiki, et al. Science 1985;230:1350-1354)被优选用于获得本发明的基因。特别是很难从文库中得到全长的 cDNA 时，可优选使用 RACE 法(RACE-cDNA 末端快速扩增法)，用于 PCR 的引物

可根据本文所公开的本发明的序列信息适当地选择，并可用常规方法合成。可用常规方法如通过凝胶电泳分离和纯化扩增的 DNA/RNA 片段。

本发明也涉及包含本发明的多核苷酸的载体，以及用本发明的载体或 Binlb 蛋白编码序列经基因工程产生的宿主细胞，以及经重组技术产生本发明所述多肽的方法。

通过常规的重组 DNA 技术(Science，1984：224：1431)，可利用本发明的多聚核苷酸序列可用来表达或生产重组的 Binlb 多肽。一般来说有以下步骤：

(1).用本发明的编码 Binlb 多肽的多核苷酸(或变异体)，或用含有该多核苷酸的重组表达载体转化或转导合适的宿主细胞；

(2).在合适的培养基中培养的宿主细胞；

(3).从培养基或细胞中分离、纯化蛋白质。

本发明中，Binlb 多核苷酸序列可插入到重组表达载体中。术语“重组表达载体”指本领域熟知的细菌质粒、噬菌体、酵母质粒、植物细胞病毒、哺乳动物细胞病毒如腺病毒、逆转录病毒或其他载体。总之，只要能在宿主体内复制和稳定，任何质粒和载体都可以用。表达载体的一个重要特征是通常含有复制起点、启动子、标记基因和翻译控制元件。

本领域的技术人员熟知的方法能用于构建含 Binlb 编码 DNA 序列和合适的转录/翻译控制信号的表达载体。这些方法包括体外重组 DNA 技术、DNA 合成技术、体内重组技术等。所述的 DNA 序列可有效连接到表达载体中的适当启动子上，以指导 mRNA 合成。这些启动子的代表性例子有：大肠杆菌的 lac 或 trp 启动子；λ 噬菌体 P_L 启动子；真核启动子包括 CMV 立即早期启动子、早期和晚期 SV40 启动子、反转录病毒的 LTRs 和其他一些已知的可控制基因在原核或真核细胞或其病毒中表达的启动子。表达载体还包括翻译起始用的核糖体结合位点和转录终止子。

此外，表达载体优选地包含一个或多个选择性标记基因，以提供用于选择转化的宿主细胞的表型性状，如真核细胞培养用的二氢叶酸还原酶、新霉素抗性以及绿色荧光蛋白(GFP)，或用于大肠杆菌的四环素或氨苄青霉素抗性。

包含上述的适当 DNA 序列以及适当启动子或者控制序列的载体，可以用于转化适当的宿主细胞，以使其能够表达蛋白质。

宿主细胞可以是原核细胞，如细菌细胞；或是低等真核细胞，如酵母细胞；或是高等真核细胞，如哺乳动物细胞。代表性例子有：大肠杆菌，链霉菌属；鼠伤寒沙门氏菌的细菌细胞；真菌细胞如酵母；植物细胞；果蝇 S2 或 Sf9 的昆虫细胞；CHO、COS、293 细胞、或 Bowes 黑素瘤细胞的动物细胞等。

本发明的多核苷酸在高等真核细胞中表达时，如果在载体中插入增强子序列时将会使转录得到增强。可举的例子包括在复制起始点晚期一侧的 100 到 270 个碱

基对的 SV40 增强子、在复制起始点晚期一侧的多瘤增强子以及腺病毒增强子等。

本领域一般技术人员都清楚如何选择适当的载体、启动子、增强子和宿主细胞。

用重组 DNA 转化宿主细胞可用本领域技术人员熟知的常规技术进行。当宿主为原核生物如大肠杆菌时，能吸收 DNA 的感受态细胞可在指数生长期后收获，用 $CaCl_2$ 法处理，所用的步骤在本领域众所周知。另一种方法是使用 $MgCl_2$。如果需要，转化也可用电穿孔的方法进行。当宿主是真核生物，可选用如下的 DNA 转染方法：磷酸钙共沉淀法，常规机械方法如显微注射、电穿孔、脂质体包装等。

获得的转化子可以用常规方法培养，表达本发明的基因所编码的多肽。根据所用的宿主细胞，培养中所用的培养基可选自各种常规培养基。在适于宿主细胞生长的条件下进行培养。当宿主细胞生长到适当的细胞密度后，用合适的方法(如温度转换或化学诱导)诱导选择的启动子，将细胞再培养一段时间。

在上面的方法中的重组多肽可在细胞内、或在细胞膜上表达、或分泌到细胞外。如果需要，可利用其物理的、化学的和其它特性通过各种分离方法分离和纯化重组的蛋白。这些方法是本领域技术人员所熟知的。这些方法的例子包括但并不限于：常规的复性处理、用蛋白沉淀剂处理(盐析方法)、离心、渗透破菌、超处理、超离心、分子筛层析(凝胶过滤)、吸附层析、离子交换层析、高效液相层析(HPLC)和其它各种液相层析技术及这些方法的结合。

重组的 Binlb 蛋白或多肽有多方面的用途。这些用途包括(但不限于)：直接做为药物治疗泌尿生殖系统感染疾病，和用于筛选促进或对抗 Binlb 蛋白功能的抗体、多肽或其它配体。用表达的重组 Binlb 蛋白筛选多肽库可用于寻找有治疗价值的能抑制或刺激 Binlb 蛋白功能的多肽分子。

另一方面，本发明还包括对 Binlb DNA 或是其片段编码的多肽具有特异性的多克隆抗体和单克隆抗体，尤其是单克隆抗体。这里，“特异性”是指抗体能结合于 Binlb 基因产物或片段。较佳地，指那些能与 Binlb 基因产物或片段结合但不识别和结合于其它非相关抗原分子的抗体。本发明中抗体包括那些能够结合并抑制 Binlb 蛋白的分子，也包括那些并不影响 Binlb 蛋白功能的抗体。本发明还包括那些能与修饰或未经修饰形式的 Binlb 基因产物结合的抗体。

本发明不仅包括完整的单克隆或多克隆抗体，而且还包括具有免疫活性的抗体片段，如 Fab'或 $(Fab)_2$ 片段；抗体重链；抗体轻链；或嵌合抗体等。

本发明的抗体可以通过本领域内技术人员已知的各种技术进行制备。例如，纯化的 Binlb 基因产物或者其具有抗原性的片段，可被施用于动物以诱导多克隆抗体的产生。与之相似的，表达 Binlb 蛋白或其具有抗原性的片段的细胞可用来免疫动物(如家兔，小鼠，大鼠等)来生产抗体。多种佐剂可用于增强免疫反应，包括但

不限于弗氏佐剂等。

本发明的单克隆抗体可以利用杂交瘤技术来制备(见 Kohler 等人, Nature 256:495, 1975; Kohler 等人, Eur. J. Immunol. 6:511, 1976; Kohler 等人, Eur. J. Immunol. 6:292, 1976; Hammerling 等人, In Monoclonal Antibodies and T Cell Hybridomas, Elsevier, N.Y., 1981)。本发明的抗体包括能阻断 Binlb 蛋白功能的抗体以及不影响 Binlb 蛋白功能的抗体。本发明的各类抗体可以利用 Binlb 基因产物的片段或功能区，通过常规免疫技术获得。这些片段或功能区还可以利用重组方法制备或利用多肽合成仪合成。与 Binlb 基因产物的未修饰形式结合的抗体可以用原核细胞(例如 *E. Coli*)中生产的基因产物来免疫动物而产生；与翻译后修饰形式结合的抗体(如糖基化或磷酸化的蛋白或多肽)，可以用真核细胞(例如酵母或昆虫细胞)中产生的基因产物来免疫动物而获得。

抗 Binlb 蛋白的抗体可用于免疫组织化学技术中，检测活检标本中的 Binlb 蛋白。此外，与 Binlb 蛋白结合的单克隆抗体也可用放射性同位素标记，注入体内可跟踪其位置和分布。

利用本发明蛋白，通过各种常规筛选方法，可筛选出与 Binlb 蛋白发生相互作用的物质，如受体、抑制剂、激动剂或拮抗剂等。

本发明蛋白及其抗体、抑制剂、激动剂、拮抗剂或受体等，当在治疗上进行施用(给药)时，可提供不同的效果。通常，可将这些物质配制于无毒的、惰性的和药学上可接受的水性载体介质中， 其中 pH 通常约为 5-8，较佳地 pH 约为 6-8，尽管 pH 值可随被配制物质的性质以及待治疗的病症而有所变化。配制好的药物组合物可以通过常规途径进行给药，其中包括(但并不限于)：肌内、静脉内 、皮下、皮内、或局部给药。

本发明的多肽可直接用于疾病治疗，例如，用于泌尿生殖系统感染方面的治疗。在使用本发明 Binlb 蛋白时，还可同时使用其他药剂，如青霉素等抗菌素。

本发明还提供了一种药物组合物，它含有安全有效量的本发明 Binlb 多肽或其激动剂、拮抗剂以及药学上可接受的载体或赋形剂。这类载体包括(但并不限于)：盐水、缓冲液、葡萄糖、水、甘油、乙醇、及其组合。药物制剂应与给药方式相匹配。本发明的药物组合物可以被制成针剂形式，例如用生理盐水或含有葡萄糖和其他辅剂的水溶液通过常规方法进行制备。诸如片剂和胶囊之类的药物组合物，可通过常规方法进行制备。药物组合物如针剂、溶液、片剂和胶囊宜在无菌条件下制造。活性成分的给药量是治疗有效量，例如每天约 1 微克/千克体重-约 5 毫克/千克体重。此外，本发明的多肽还可与其他治疗剂一起使用。

使用药物组合物时，是将安全有效量的 Binlb 蛋白或其拮抗剂、激动剂施用于哺乳动物，其中该安全有效量通常至少约 10 微克/千克体重，而且在大多数情况

下不超过约 8 毫克/千克体重，较佳地该剂量是约 10 微克/千克体重-约 1 毫克/千克体重。当然，具体剂量还应考虑给药途径、病人健康状况等因素，这些都是熟练医师技能范围之内的。此外，Binlb 核苷酸序列也可用于多种治疗目的。

能与 Binlb 蛋白结合的多肽分子可通过筛选由各种可能组合的氨基酸结合于固相物组成的随机多肽库而获得。

本发明还涉及定量和定位检测 Binlb 蛋白水平的诊断试验方法。这些试验是本领域所熟知的，且包括 FISH 测定和放射免疫测定。试验中所检测的 Binlb 蛋白水平，可以用作解释 Binlb 蛋白在各种疾病中的重要性和用于诊断 Binlb 蛋白起作用的疾病。

一种检测检测样品中是否存在Binlb蛋白的方法是利用Binlb蛋白的特异性抗体进行检测，它包括：将样品与Binlb蛋白特异性抗体接触；观察是否形成抗体复合物，形成了抗体复合物就表示样品中存在Binlb蛋白。

Binlb 核苷酸序列可用于 Binlb 蛋白相关疾病的诊断和治疗。在诊断方面，Binlb 核苷酸序列可用于检测 Binlb 蛋白的表达与否或在疾病状态下 Binlb 蛋白的异常表达。如 Binlb DNA 序列可用于对活检标本的杂交以判断 Binlb 蛋白的表达异常。杂交技术包括 Southern 印迹法，Northern 印迹法、原位杂交等。这些技术方法都是公开的成熟技术，相关的试剂盒都可从商业途径得到。本发明的多核苷酸的一部分或全部可作为探针固定在微阵列(microarray)或 DNA 芯片(又称为“基因芯片”)上，用于分析组织中基因的差异表达分析和基因诊断。用 Binlb 蛋白特异的引物进行 RNA-聚合酶链反应(RT-PCR)体外扩增也可检测 Binlb 蛋白的转录产物。

检测 Binlb 基因的突变也可用于诊断 Binlb 蛋白相关的疾病。Binlb 蛋白突变的形式包括与正常野生型 Binlb DNA 序列相比的点突变、易位、缺失、重组和其它任何异常等。可用已有的技术如 Southern 印迹法、DNA 序列分析、PCR 和原位杂交检测突变。另外，突变有可能影响蛋白的表达，因此用 Northern 印迹法、Western 印迹法可间接判断基因有无突变。

本发明的序列对染色体鉴定也是有价值的。简而言之，根据本发明Binlb蛋白的cDNA制备PCR引物(优选15-35bp)，可以将序列定位于染色体上。然后，将这些引物用于PCR筛选含各条人染色体的体细胞杂合细胞。只有那些含有相应于引物的人基因的杂合细胞会产生扩增的片段。

一旦序列被定位到准确的染色体位置，此序列在染色体上的物理位置就可以与基因图数据相关联。这些数据可见于例如，V. Mckusick, Mendelian Inheritance in Man(可通过与Johns Hopkins University Welch Medical Library联机获得)。然后可通过连锁分析，确定基因与业已定位到染色体区域上的疾病之间的关系。

本发明的 Bin1b 为治疗泌尿生殖系统感染等疾病以及开辟男性避孕方法提供新的途径，因而具有巨大的应用前景。

下面结合具体实施例，进一步阐述本发明。应理解，这些实施例仅用于说明本发明而不用于限制本发明的范围。下列实施例中未注明具体条件的实验方法，通常按照常规条件如Sambrook等人， 分子克隆：实验室手册(New York: Cold Spring Harbor Laboratory Press, 1989)中所述的条件，或按照制造厂商所建议的条件。

实施例 1

Bin1b mRNA 的差异片段的发现：

基本上按 Liang, P.等在“Science”上发表的方法[P. Liang, and A. B. Pardee, *Science* 257, 967 (1992)]，从 Sprague-Dawley 大鼠附睾的头部、体部、尾部分别抽提总 RNA，并用不含 RNase 的 DNase 消化去除留存的染色体 DNA。经沉淀后，用 2 ug 分别来自头、体、尾部的纯化的总 RNA 进行反转录，反转录反应中，使用 2.5 uM 的下游引物 T11CA (5'-TTTTTTTTTTTCA-3')和 400 单位的 MMLV 反转录酶。以 1/20 的反转录产物为模板，进行 PCR 扩增，在 20 ul 反转录体系中，包含 2.5 uM 的上游引物 502 (5'-TGGATTGGTC-3')，0.5 uM 的下游引物 T11CA，10 mM Tris-HCl pH 9.0， 1.5 mM $MgCl_2$， 50 mM KCl， 0.1% Triton X-100， 4 μM each dNTP， 1 μCi ^{32}P-dATP 和 3 单位 Taq 聚合酶。反应条件为 94 ℃ 5 min；然后是 40 个循环： 94 ℃ 30 秒，40 ℃ 60 秒，72 ℃ 50 秒；最后是 72 ℃ 10 min。PCR 反应产物经沉淀后，在 0.2 mm 厚的 6%的测序胶上分离。

在大鼠附睾头部有一条特异表达 340bp 的条带(命名为 Bin1，如图 1 中箭头所示)。将该条带割出后，于沸水中洗脱 15 分钟。被洗脱出的 DNA 经乙醇沉淀后，用上面相同的条件进行 PCR 扩增，但需改变 dNTP 的浓度为 40 uM 并且不加同位素。扩增出来的 DNA 片段克隆于 pBluescript SK$^-$ 质粒中。用反向 Northern 方法从 52 个克隆中筛选出在附睾头部有特异杂交信号的克隆，并命名为 Bin1b。在这个基因片段基础上，进一步采用 5'-RACE 的方法克隆该基因的全长 cDNA。

实施例 2

Bin1b 全长 cDNA 的克隆和结构分析：

根据已有的 cDNA 片段设计引物，通过类似的 RT-PCR 反应向 5'-末端延伸。使用了两种 5'-RACE 的方法获得 cDNA 全长。

第一种方法参照 Apte, A. N.的报道[A. N. Apte, P. D. Siebert, in *Reverse Transcriptase PCR* J. W. Larrick, P. D. Siebert, Eds. (Ellis Horwood, London, 1995) pp. 232-244.]，利用 DNA oligo 和 cDNA 第一链连接的方式。

另外一种方法修订于 Maruyama, K. [K. Maruyama and S. Sugano, *Gene* 138, 171(1994)]的报道，我们做了一定的修改，这种方法克服了前一种方法的缺陷，可以确保获得 cDNA 的 5'-帽子位点。它使用 DNA oligo 代替 RNA oligo 和总 RNA 进行连接。具体过程如下：将 50 ug 大鼠附睾头部的总 RNA 用 400 单位的细菌碱性磷酸酶(Bacterial Alkaline phosphatase, BAP)于 37°C 消化 30 分钟，随后 65°C 消化 30 分钟。接着，用 50 ng/ul 的蛋白酶 K 于 37°C 酶解 30 分钟，以便消化 BAP。经纯化后，将其中 10 ug RNA 用 2 单位的烟草酸焦磷酸酶(Tobacco Acid Pyrophosphatase, TAP)于 37°C 酶解 2 小时，经苯酚/氯仿抽提后，乙醇沉淀。分别取 0.75 ug (3 pmol)未经任何处理的 RNA、BAP 处理的 RNA、和 TAP 处理的 RNA 与 1.25 pmol DNA oligo #7209 (5'-AATGGTACCGTGACGTGGTCC-3') (SEQ ID NO: 5) 在 10 ul 反应体系中于 17 °C 连接 18 小时，连接反应体系中包括 1.2 单位/ul T4 RNA 连接酶，50 mM Tris-HCl (pH 8.0)，10 mM $MgCl_2$，1 mM 氯化六氨合高钴(hexammine) colbalt chloride)，25% PEG8000，1 mM ATP。1/50 的连接产物被用来进行 RT-PCR 反应，反应体系采用 GIBCO BRL 公司的 SuperScript II One-Step RT-PCR System，反应体积为 20 ul。另外，还包括 200 nM 的 Binlb 基因特异的引物(GSP) (5'- TGGCCCCGCTGCATGAAGCAC-3') (SEQ ID NO: 6)和 200 nM 的 DNA oligo #7209 (5'-AATGGTACCGTGACGTGGTCC-3') (SEQ ID NO: 5)。反应条件为：50 °C 30 min，94 °C 2 min， 然后按照以下条件进行 35 个循环，94 °C 5 秒，60 °C 15 秒，72 °C 45 秒， 最后于 72 °C 延伸 5 min。取 0.2 ul RT-PCR 反应产物作为模板进行第二轮 PCR 反应，反应体积为 10 ul，反应体系中含有 50 mM Tris-HCl pH 8.3, 1-3mM $MgCl_2$, 250 μg/ml BSA, 0.5% Ficoll 400, 1mM tartrazine, 200 μM dNTP, 500 nM Binlb GSP, 500 nM DNA oligo #7209, 和 0.4 单位的 Taq 聚合酶。反应在毛细管中进行，反应条件为：94 °C 1 min，然后按下列条件进行 60 个循环：94 °C 0 秒，60 °C 0 秒，77 °C 15 秒，最后于 77 °C 延伸 5 min。PCR 产物克隆于 pBluescript SK* T-载体中，并进行测序。这样，本发明人就获得了 Binlb 的 3'和 5'两个 cDNA 片段。

在 2 个 cDNA 片段的两端分别设计引物(5'-GGACACCCAG TCATCAGTCA-3' (SEQ ID NO: 7)和 5'-CAGCAAGTGT TTATTGAGCA-3' (SEQ ID NO: 8))，使用 RT-PCR 反应产物作为模板，通过 PCR 反应获得 Binlb 的全长 cDNA 克隆。这个序列在 6 只不同龄的大鼠中得到了验证。Binlb 的全长 cDNA 为 385bp (SEQ ID NO: 1) (如图 2A 所示)。它编码一个 68 个氨基酸的多肽(命名为 Binlb 蛋白)(SEQ ID NO: 2)，在其 N 端有一个 16 个氨基酸的信号肽，它的 C-端 45 个氨基酸为理论上的成熟多肽(另有 7 个氨基酸在 pro-Binlb 转变为成熟蛋白时被切除)。

通过 Genebank 的 Blastn 程序查询发现，Binlb 在编码区 180-241 位与人精子抗原 HE2(673bp)的非编码区(457-518 位)有 83%的相似性。此外，通过 Genebank

的 Blastp 程序查询发现 Binlb 编码的多肽与哺乳动物β-防卫素家族有一定程度的相似性(如图 2B 所示)。

继而，本发明人用类似的 PCR 方法克隆了 Binlb 的基因组 DNA(SEQ ID NO: 4)(如图 2A 所示)。证明了 Binlb 基因结构也附合β-defensin 家族的规律，即由一个内含子和两个外显子组成。

目前，关于 HE2 的研究发展很快，有两家实验室相继报导了人 HE2 和大猩猩 HE2(EP2)的多个异构体。有一些异构体虽然它们的表达量很低，然而与 Binlb 有非常好的同源性(如图 2C 所示)。

实施例 3

Binlb 蛋白的表达

在该实施例中，通过体外转录和翻译实验，验证了 Binlb 表达出多肽的大小与理论上推断的一致，约为 31kDa。

将带有约 75bp polyA 尾部且包括编码仅 68 个氨基酸(7799 道尔顿)的 ORF 的 Binlb 全长 cDNA，克隆在 pSPT18(Promega)的 EcoRI 和 HindIII 之间，位于 T7 启动子下游，形成质粒 pSPT18-Binlb(图 3C)。用 HindIII 线性化的质粒，在含有放射性标记的 ^{35}S-Met 的 TNT T7 Coupled Reticulocyte Lysate System 中分析，并用 16.5%Tricine-SDS-PAGE 凝胶进行分析(按 Schagger 和 Von Jagow 所述的方法)，结果如图 3A 所示。同时，使用阴性对照(无质粒 DNA)和荧光素酶 SP6 和 T7 对照进行比较(图 3B)。

实施例 4

Binlb 的融合表达及抗体制备

将 Binlb cDNA 的 126-281 位的片段接入融合表达载体 pQE-40(Qiagen 公司)中。挑选出的阳性克隆，经 1mM IPTG 诱导表达后，用 Ni-NTA-琼脂糖(Qiagen 公司)纯化。该实验的具体操作过程详见 QiaExpressionist(Qiagen 公司)。

纯化后的融合表达蛋白 DHFR-Binlb 的大小为 31KD(如图 4 所示)。将纯化的蛋白免疫兔子，取得抗 DHFR-1b 的抗血清。然而，该抗血清的抗 Binlb 的滴度不高，有待进一步改进。

实施例 5

Binlb 表达的组织分布

Binlb 表达的组织分布的 Northern 印迹分析，按 1984 年 Church 和 Gilbert 首先报道的方法进行。从 Sprague-Dawley 大鼠附睾的头部、睾丸、前列腺、肝、心、脾、肺、肾、胸腺、肾上腺、凝结腺(Coagulating gland)、精囊、下丘脑、纹状体、垂体、海马、小脑、大脑皮层中分别抽提总 RNA。各取 20 ug 总 RNA，

在 1.2% 的甲醛琼脂糖电泳上分离，并转膜 Hybond™-N+尼龙滤膜(Amersham Pharmacia Biotech)。Binlb 的探针用 Promega 公司的 Prime-a-Gene® System 试剂盒标记。杂交结果表明，Binlb 只在大鼠附睾头部有表达(如图 5A 所示)。这也促使我们用原位杂交来更精确地定位 Binlb 的表达特异性。

实施例 6

Binlb 表达在附睾中的定位

在该实施例中，用原位杂交法对 Binlb 进行定位，方法如下：

(1)45 天龄大鼠附睾，以 4%多聚甲醛室温固定 4 小时以上，固定后的组织经常规梯度乙醇脱水、二甲苯透明、56-58 ℃高温石蜡包埋，常规切片，切片厚 10 微米，贴于硅烷处理的载玻片上，42 ℃烤片 48-72 小时，密封于玻璃染色缸内，保存于 4 ℃。

(2)组织切片经二甲苯脱蜡、梯度乙醇复水至 DEPC-H_2O 中，于 0.2 M HCl 中处理 20 分钟，DEPC-H_2O 洗 5 分钟，2 X SSC 70 ℃ 30 分钟，DEPC-H_2O 洗 5 分钟，切片经 TE (pH 7.5)37 ℃ 平衡 30 分钟后，置于含 3 ug/ml 的 蛋白酶 K 的 TE(pH 7.5)中，37 ℃ 消化 10 分钟。以含 0.2 %甘氨酸的 PBS 和 PBS 溶液各洗 2 次，每次 1 分钟，再以 4%多聚甲醛室温后固定 20 分钟，之后浸于 1.5%乙酸酐/0.25 M 三乙醇胺中，室温 10 分钟，PBS 洗 5 分钟。

(3)用餐巾纸吸去载玻片上多余的水分，将预杂交液[每 5 ml 杂交 A 液中含有 0.65 ml DEPC-H2O，2.5 ml 甲酰胺，1.25 ml 20X 杂交盐液(3 M NaCl，0.1 M PIPES，0.1 M EDTA pH 6.8)，0.5 ml 100X Denhardt's，0.1 ml 10%SDS。将 125 ul 鱼精 DNA/酵母 tRNA(10 mg/ml)煮沸 5 分钟，立即冷却于碎冰中，加入 4.875 ml 杂交 A 液，混合后，其中 4 ml 作为预杂交液，1 ml 留作制备杂交液。]滴在组织切片上， 然后水平放置于湿盒内，60 ℃ 4 小时。

(4)取适量 Dig 标记的反义 RNA 探针和有义 RNA 探针(探针的标记参照 Dig RNA Labeling Kit，宝灵曼公司)，分别加入等量的甲酰胺，煮沸 30 秒，立即冷却于碎冰中，加入适量预杂交液，充分混合即为杂交液。用吸水纸吸去切片上的预杂交液，将杂交液滴加在组织切片上，每张切片 40 ul，其中含 1.2 ug 探针，探针浓度为 30 ng/ul(可作不同的浓度，以滴定出最佳浓度)。同时做抗体对照(不加探针，加抗体)和空白对照(探针、抗体都不加)。在杂交液上加盖玻片，切片水平放置在密封湿盒内，60 ℃ 杂交过夜。

(5)将载玻片置 4XSSC 中去掉盖玻片，再以 4XSSC 洗四次，每次 5 分钟，于含有 20 ug/ml RNase A 的 RNase 缓冲液(含 0.5 M NaCl 的 TE)中 37℃ 消化 30 分钟，再于 RNase 缓冲液中 30℃ 吸 30 分钟，于 2XSSC 中 50 ℃ 洗 2 次，每次 15 分钟，于 0.1XSSC 中 50 ℃ 洗 2 次，每次 15 分钟。

(6)切片置 TBS(100 mM Tris-HCl pH 7.5, 150 mM NaCl)中洗 5 分钟，将含

有 0.5%封闭剂(宝灵曼公司)的封闭液滴在组织切片上，室温封闭 1 小时，吸去封闭液，滴加 1：1000(可滴定)封闭液稀释的 Anti-Dig-AP 抗体，室温结合 90 分钟或 4 ℃过夜，TBS 洗 4 次，每次 5 分钟。切片置于显色缓冲液 CDS(100 mM Tris-HCl pH 9.5, 100 mM NaCl, 50 mM $MgCl_2$)中 5 分钟，将显色液[每 ml 含 4.5 ul 氮蓝四唑(nitroblue tetrazolium(NBT), 75 mg/ml), 3.5 ul 5-溴-4-氯-3-吲哚-磷酸 (5-bromo-4-chloro-3-indolyl-phosphate(BCIP), 50 mg/ml), 2 ul 1M Levamisole, CDS 990 ul]滴于组织切片上，置于避光湿盒内，37 ℃显色 1-5 小时或室温显色过夜。显色后以 TE pH 8.0 洗 3 次，每次 5 分钟，终止反应。再以双蒸水洗 5 分钟，梯度乙醇脱水，二甲苯透明，中性树胶封片。经显微镜观察，Binlb 定位在大鼠附睾头部中央的上皮细胞中(如图 5B-E 所示)。

实施例 7

Binlb 在大鼠发育过程中的变化

Binlb 在大鼠发育过程中的变化的 Northern 印迹分析方法同实施例 5，RNA 样品提取自 15 天龄大鼠的附睾头、30、45、60、120、270、720 天龄大鼠的附睾头、附睾体和附睾尾。

结果表明，Binlb 的表达水平在性成熟期(包括性活跃期)达到最高，然后逐渐下降(如图 5F 所示)。这提示 Binlb 与男性精子成熟有关，并可能可以调控 Binlb 进而实现男性避孕。

实施例 8

Binlb 的表达受激素等小分子的调控

在该实施例中，通过 EDS 模型观察到 Binlb 的表达至少是部分受雄激素正调控的。

用二甲亚砜-水(1:3, v/v)中的 EDS(7.5 mg /100 g 体重)，对成年 Sprague-Dawley 大鼠进行腹膜内注射一次。在之后间隔 1 天，3 天，7 天，14 天，21 天，28 天，42 天等不同时间，将大鼠杀死。另一组未注射 EDS 的大鼠也被杀死，作为空白对照。从大鼠附睾头部区域抽提的总 RNA(20 微克)，用 1.2%甲醛-琼脂糖凝胶电泳分级，印迹于 Hybond+尼龙滤膜，然后再依次与 Binlb 3'端 cDNA 探针和 18s 核糖体 RNA 探针杂交。

在注射 EDS 后，雄激素的分泌会受抑制而下降，但约 2 周后雄激素的分泌又会恢复。与雄激素水平的变化相比，可以看出 Binlb 的表达受雄激素的正调控(图 6A 和 6B)。

这提示，可以通过激素小分子来影响 Binlb 基因的表达，从而调控精子的成熟，进而为设计男性避孕药开辟了新途径。

实施例 9

Binlb 是一个天然抗菌肽

上述的一系列研究提示，Binlb 是β-防卫素家族的一个成员且与精子成熟相关。在该实施例中，证实了 Binlb 的抗菌活性。

为了验证 Binlb 杀菌功能，检验了大鼠附睾头部细胞培养的分泌物的杀菌功能。100 CFU(colony forming units)的大肠杆菌(E. coli)与大鼠附睾头部和尾部的上皮细胞分别共培养。16 个小时以后，收集细胞的培养液，用来检测杀菌活力，发现头部上皮细胞的培养液有很强的杀菌活力而尾部的则无(如图 7A 所示)。这首先肯定了附睾头部细胞的分泌物有杀菌活力。

虽然 Binlb 表达在附睾的头部，但为了确定杀菌活力是否属于 Binlb 蛋白，本发明人在与大肠杆菌共培养之前，先加入 Binlb 的反义 RNA，用以阻断 Binlb 的表达。结果发现，Binlb 的反义 RNA 来抑制 Binlb 蛋白的产生后，附睾头部细胞的分泌物的杀菌活力随之消失，从而证实了 Binlb 的抗菌性。(如图 7B 所示)。

综上所述，Binlb 蛋白为一未见报道的属于β-防卫素家族的新型天然抗菌肽。

实施例 10

Binlb 的表达受炎症上调

将大鼠输精管结扎，引起附睾中精子堆积而发生炎症感染时，附睾头部 Binlb mRNA 比结扎前增加约 3 倍，而在尾部则无任何改变(如图 7C)。这表明，Binlb 的表达受炎症上调。

在本发明提及的所有文献都在本申请中引用作为参考，就如同每一篇文献被单独引用作为参考那样。此外应理解，在阅读了本发明的上述讲授内容之后，本领域技术人员可以对本发明作各种改动或修改，这些等价形式同样落于本申请所附权利要求书所限定的范围。

序 列 表

(1)一般信息:

(ii)发明名称: 新的天然抗菌肽、其编码序列及用途

(iii)序列数目:6

(2)SEQ ID NO: 1的信息:

(i)序列特征:

(A)长度: 385bp

(B)类型: 核酸

(C)链性: 双链

(D)拓扑结构: 线性

(ii)分子类型: cDNA

(xi)序列描述: SEQ ID NO: 1:

GGACACCCAG TCATCAGTCA CATCTGCTTT CCTGCACAGA GAGAGCGCCA TAAAACATGA 60
AGGTTTTGTT ACTCTTTGCT GTTTTCTTCT GCTTGGTCCA AAGAAACTCA GGGGACATAC 120
CACCTGGAAT CAGAAACACC GTGTGCTTCA TGCAGCGGGG CCACTGTAGG CTCTTCATGT 180
GCCGTTCTGG GGAGAGAAAG GGGGATATTT GCTCTGACCC CTGGAACAGA TGCTGCGTAT 240
CCAGTTCCAT TAAAAACAGA TGATAGAAGA CTCATTGGAA GATCTGAGAT GTGGGGTGCA 300
AGCTCTTGGA AGCTAGAGAC CTGGAAGCAC CCCAAAGGCT TTGAGTATGT GTGGCTAATG 360
GTGCGTGCTC AATAAACACT TGCTG 385

(2)SEQ ID NO: 2的信息:

(i)序列特征:

(A)长度: 68个氨基酸

(B)类型: 氨基酸

(D)拓扑结构: 线性

(ii)分子类型: 多肽

(xi)序列描述: SEQ ID NO: 2:

MKVLLLFAVF FCLVQRNSGD IPPGIRNTVC FMQRGHCRLF MCRSGERKGD 50
ICSDPWNRCC VSSSIKNR 68

(2)SEQ ID NO: 3的信息:

(i)序列特征:

(A)长度: 45个氨基酸

(B)类型: 氨基酸

(D)拓扑结构: 线性

(ii)分子类型：多肽

(xi)序列描述：SEQ ID NO: 3:

GIRNTVCFMQ RGHCRLFMCR SGERKGDICS DPWNRCCVSS SIKNR 45

(2)SEQ ID NO:4 的信息

(i)序列特征

(A)长度：1696碱基

(B)类型：核酸

(C)链性：双链

(D)拓扑结构：线性

(ii)分子类型：DNA(基因组)

(xi)序列描述：SEQ ID NO：4:

GGACACCCAG TCATCAGTCA CATCTGCTTT CCTGCACAGA GAGAGCGCCA TAAAACATGA 60

AGGTTTTGTT ACTCTTTGCT GTTTTCTTCT GCTTGGTCCA AAGAAACTCA Ggtaaatgtc 120

ttctgagtag ccctggagaa ggcaggatgc ccttttaggt ttgtagacca cattgaggtg 180

tgtccaggta tcaacattgg gcacagatgg tgggccactc tggggctcag ggtcggacca 240

ctttcctaac gaagaggttt tattttgatt ttttttgtt tgttcatttg tcaagagttg 300

caaattttac agcacggaga cacagaggcc tatattctcc attgtgaata agaaggtctg 360

attgtaactt gagagtttat tcaggacaga attacagccg tacctgtgtc aaaagtgtaa 420

ttttactgcc tcgctgtgag cagagaaggt gttcacattt atgcccccttc cctacccatt 480

acatccacag aacaccagat gtatgcttta aatgaatttt caaatgagag aaaaataggt 540

tcctttaaga aagctagagt ccaggtcctg aagccttgaa ttgctggcag ttctgtcaag 600

gtggactaca cccacatctc catgaacctt cccaaccatg gtaaaccgga tgaacacagt 660

atcacaaatc agtccccagc tgaagtccgg ctattgcagg agaccagttt cctaaatgtt 720

acaggcatag gttgggccgc tgttgctttt taacacaggg tgtgcaacat tgttaaaaag 780

gtttttttta accatctctt tcccatggtg ctttcttttg ggggactcta gttgtttttg 840

ttttgtttta ttgttttact tagaaggaca cacaagacac attgttatct ttcttcttct 900

tattgtagtc ataagagtga aaacccaacc atgagctgag acagacccgc tcctaacttt 960

tctatggcct gagacccagc tcctgttatt ctgttctgtt ttctttttct tttttaattt 1020

atttatttta tgtatgtgag tacagtgtca ctgtcttcag acacaccaga agagggtgtc 1080

agatcccatt acagatggtt gtgagccacc atgtggttgc tgggaattga actcgggacc 1140

tctggaagag cagtcagtgc tcttaaccgc tgagccatct ttccagcccc tgttctgttt 1200

tcttaaatac cactccccca ctccacaatg tacctctatc tctgggcagc tgcagagccc 1260

tggcctgcaa tgggctaggt gacttcacac tcagtctgtc atgccatccc cgaaacacca 1320

cgagatataa atggttgcta ttgaaagcta aggaggaaaa tctcagtgac gccgaaactc 1380

tggaagagtg gagcagattc ttcgagaggg gctgggggct gggggctggg ggctggagcc 1440

actgttttat ctcagtctgt tgtttccaca gGGGACATAC CACCTGGAAT CAGAAACACC 1500

GTGTGCTTCA TGCAGCGGGG CCACTGTAGG CTCTTCATGT GCCGTTCTGG GGAGAGAAAG 1560

GGGGATATTT GCTCTGACCC CTGGAACAGA TGCTGCGTAT CCAGTTCCAT TAAAAACAGA 1620

TGATAGAAGA CTCATTGGAA GATCTGAGAT GTGGGGTGCA AGCTCTTGGA AGCTAGAGAC 1680
CTGGAAGCAC CCCAAA 1696

(2)SEQ ID NO: 5的信息
(i)序列特征
(A)长度：21碱基
(B)类型：核酸
(C)链性：单链
(D)拓扑结构：线性
(ii)分子类型：寡核苷酸
(xi)序列描述：SEQ ID NO：5:
AATGGTACCG TGACGTGGTC C 21

(2)SEQ ID NO:6 的信息
(i)序列特征
(A)长度：21碱基
(B)类型：核酸
(C)链性：单链
(D)拓扑结构：线性
(ii)分子类型：寡核苷酸
(xi)序列描述：SEQ ID NO：6:
TGGCCCCGCT GCATGAAGCA C 21

(2)SEQ ID NO:7 的信息
(i)序列特征
(A)长度：20碱基
(B)类型：核酸
(C)链性：单链
(D)拓扑结构：线性
(ii)分子类型：寡核苷酸
(xi)序列描述：SEQ ID NO：7:
GGACACCCAG TCATCAGTCA 20

(2)SEQ ID NO:8的信息
(i)序列特征
(A)长度：20碱基
(B)类型：核酸
(C)链性：单链
(D)拓扑结构：线性

(ii)分子类型：寡核苷酸

(xi)序列描述：SEQ ID NO：8:

CAGCAAGTGT TTATTGAGCA 20

(2)SEQ ID NO:9 的信息

(i)序列特征

(A)长度：23碱基

(B)类型：核酸

(C)链性：单链

(D)拓扑结构：线性

(ii)分子类型：寡核苷酸

(xi)序列描述：SEQ ID NO：9:

GGACACCCAG TCATCAGTCA CAT 23

(2)SEQ ID NO:10的信息

(i)序列特征

(A)长度：22碱基

(B)类型：核酸

(C)链性：单链

(D)拓扑结构：线性

(ii)分子类型：寡核苷酸

(xi)序列描述：SEQ ID NO：10:

TTTGGGGTGC TTCCAGGTCT CT 22

说 明 书 附 图

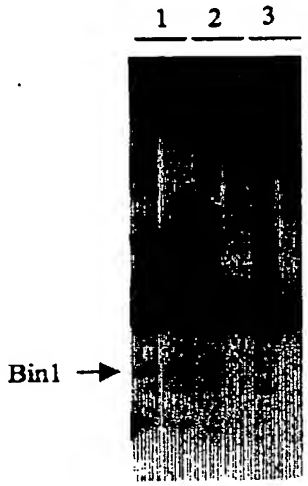

1 2 3
Bin1

图 1

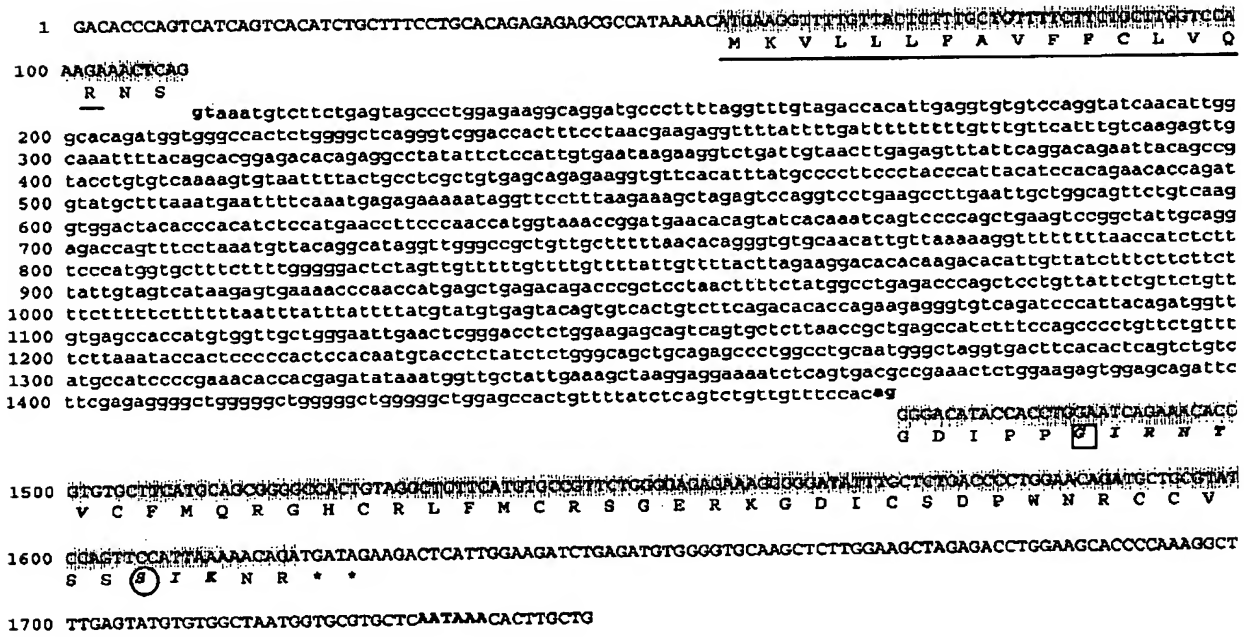

1 GACACCCAGTCATCAGTCACATCTGCTTTCCTGCACAGAGAGAGCGCCATAAAACA
M K V L L L F A V F F C L V Q
100 AAGAAACTCAG
R N S
gtaaatgtcttctgagtagccctggagaaggcaggatgccctttaggtttgtagaccacattgaggtgtgtccaggtatcaacattgg
200 gcacagatggtgggccactctggggctcagggtcggaccactttcctaacgaagaggttttatttgattttttttgtttgttcatttgtcaagagttg
300 caaattttacagcacggagacacagaggcctatattctccattgtgaataagaaggtctgattgtaacttgagagtttattcaggacagaattacagccg
400 tacctgtgtcaaaagtgtaattttactgcctcgctgtgagcagagaaggtgttcacatttatgccccttccctacccattacatccacagaacaccagat
500 gtatgctttaaatgaattttcaaatgagagaaaaataggttcctttaagaaagctagagtccaggtcctgaagccttgaattgctggcagttctgtcaag
600 gtggactacacccacatctccatgaaccttcccaaccatggtaaaccggatgaacacagtatcacaaatcagtccccagctgaagtccggctattgcagg
700 agaccagtttcctaaatgttacaggcataggttgggccgctgttgctttttaacacagggtgtgcaacattgttaaaaaggttttttttaaccatctctt
800 tcccatggtgctttctttgggggactctagttgttttgttttgttttattgtttacttagaaggacacacaagacacattgttatctttcttcttct
900 tattgtagtcataagagtgaaaacccaaccatgagctgagacagacccgctcctaacttttctatggcctgagacccagctcctgttattctgttctgtt
1000 ttctttttctttttttaatttatttatttttatgtatgtgagtacagtgtcactgtcttcagacacaccagaagagggtgtcagatcccattacagatggtt
1100 gtgagccaccatgtggttgctgggaattgaactcgggacctctggaagagcagtcagtgctcttaaccgctgagccatctttccagcccctgttctgttt
1200 tcttaaataccactcccccactccacaatgtacctctatctctgggcagctgcagagccctggcctgcaatgggctaggtgacttcacactcagtctgtc
1300 atgccatccccgaaacaccacgagatataaatggttgctattgaaagctaaggaggaaaatctcagtgacgccgaaactctggaagagtggagcagattc
1400 ttcgagaggggctgggggctgggggctgggggctggagccactgttttatctcagtctgttgtttccacag
GGGACATACCACCTGGAATCAGAAACACC
G D I P P G I R N T
1500 GTGTGCTTTCATGCAGCGGGGCCACTGTAGGCTCTTCATGTGCCGTTCTGGGCAGAGAAACGGGGATATTTAGCTCTGACCCCTGGAACAGATGCTGCGTAT
V C F M Q R G H C R L F M C R S G E R K G D I C S D P W N R C C V
1600 CCAGTTCCATTTAAAAACAGATGATAGAAGACTCATTGGAAGATCTGAGATGTGGGGTGCAAGCTCTTGGAAGCTAGAGACCTGGAAGCACCCCAAAGGCT
S S S I K N R * *
1700 TTGAGTATGTGTGGCTAATGGTGCGTGCTCAATAAACACTTGCTG

图 2A

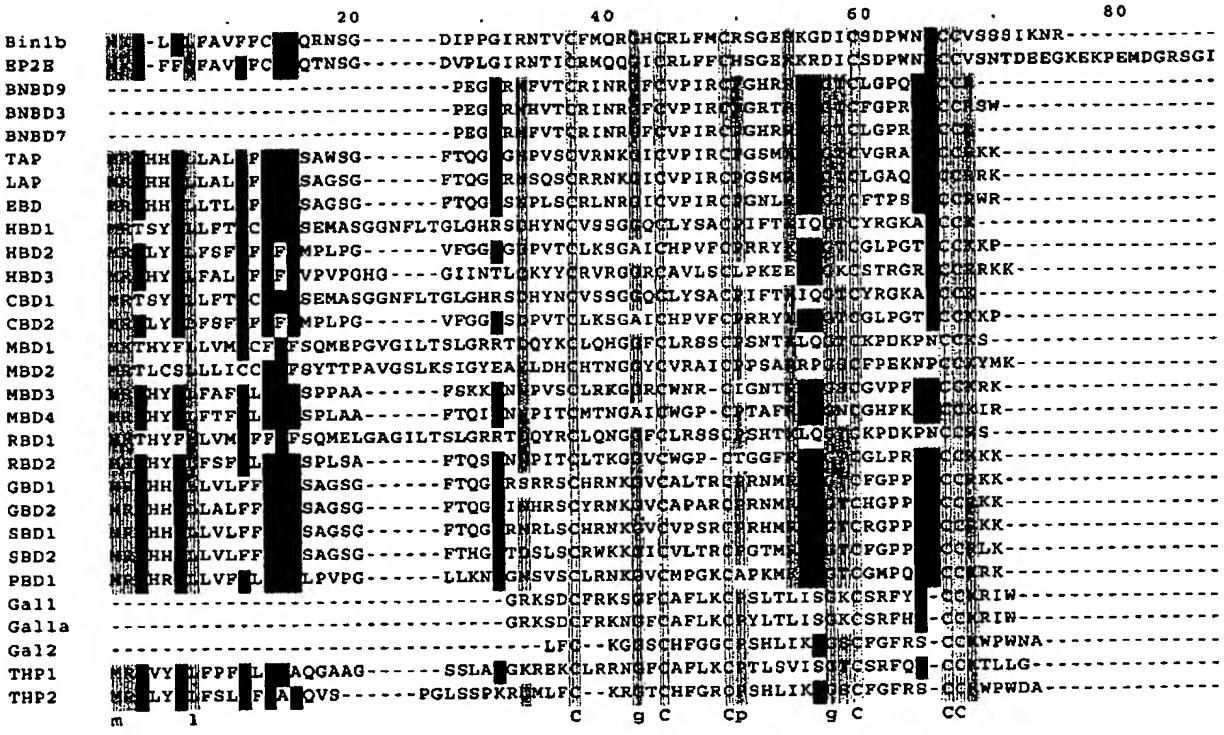

20
40
60
80
Bin1b
EP2E
BNBD9
BNBD3
BNBD7
TAP
LAP
EBD
HBD1
HBD2
HBD3
CBD1
CBD2
MBD1
MBD2
MBD3
MBD4
RBD1
RBD2
GBD1
GBD2
SBD1
SBD2
PBD1
Gal1
Gal1a
Gal2
THP1
THP2

图 2B

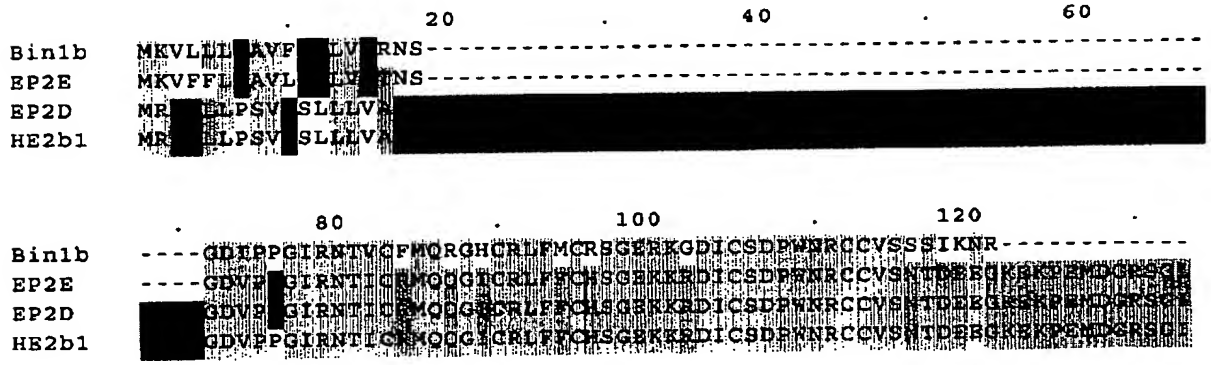
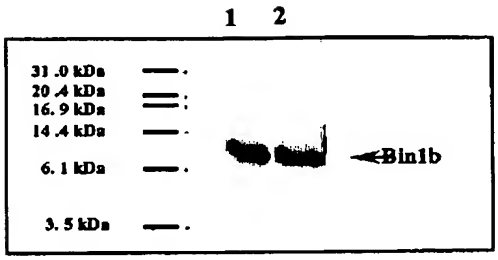
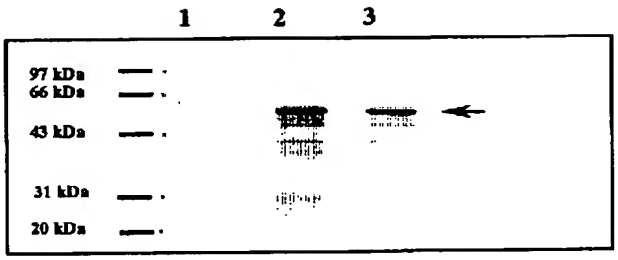

20
40
60
Bin1b
EP2E
EP2D
HE2b1
80
100
120
Bin1b
EP2E
EP2D
HE2b1

图 2C

A.



B.



C.

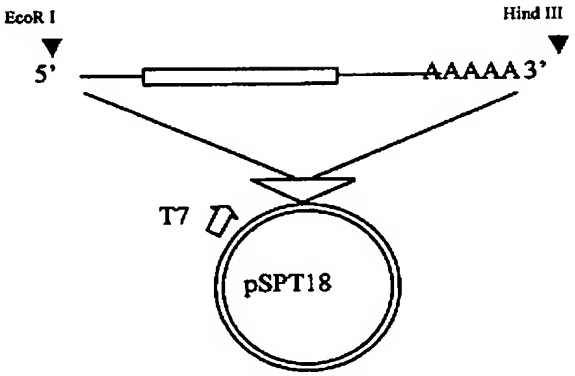


图 3

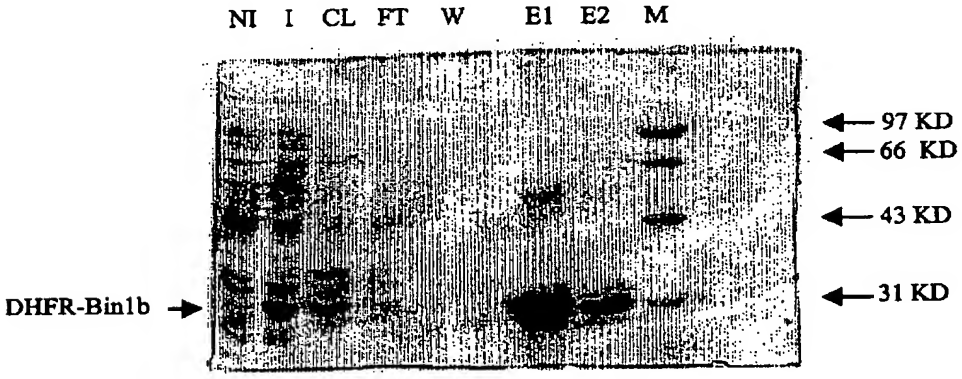


图 4

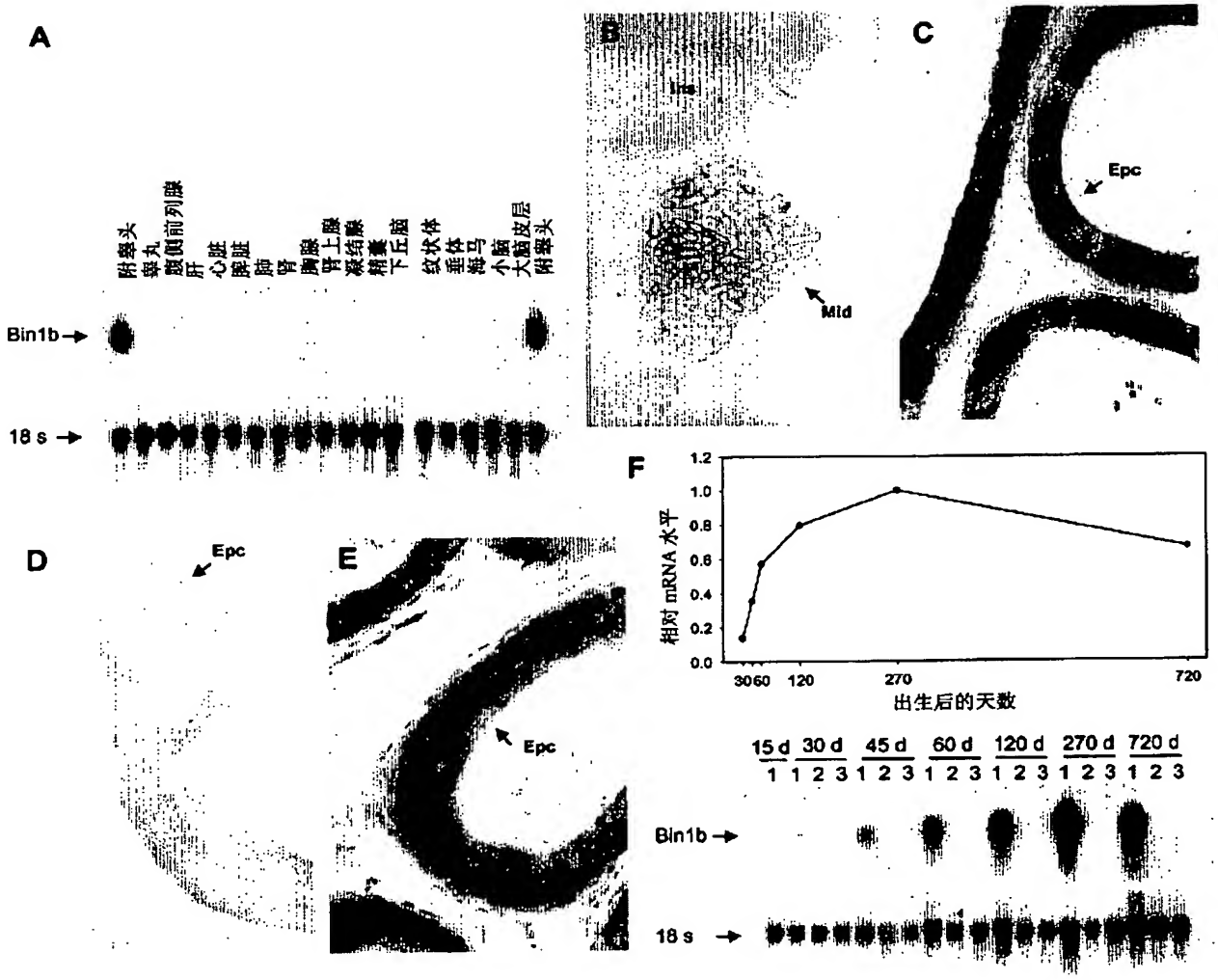
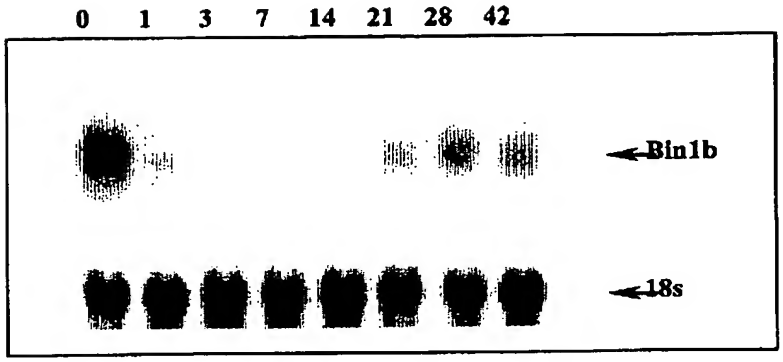

A
附睾头
睾丸
腹侧前列腺
肝
心脏
脾脏
肺
胸腺
肾上腺
凝结腺
精囊
下丘脑
纹状体
垂体
延髓
小脑
大脑皮层
附睾头
Bin1b
18 s
B
Mid
C
Epc
D
Epc
E
Epc
F
相对 mRNA 水平
1.2
1.0
0.8
0.6
0.4
0.2
0.0
3060
120
270
720
出生后的天数
15 d 30 d 45 d 60 d 120 d 270 d 720 d
1 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3
Bin1b
18 s

图 5

A.


0 1 3 7 14 21 28 42
Bin1b
18s

B.

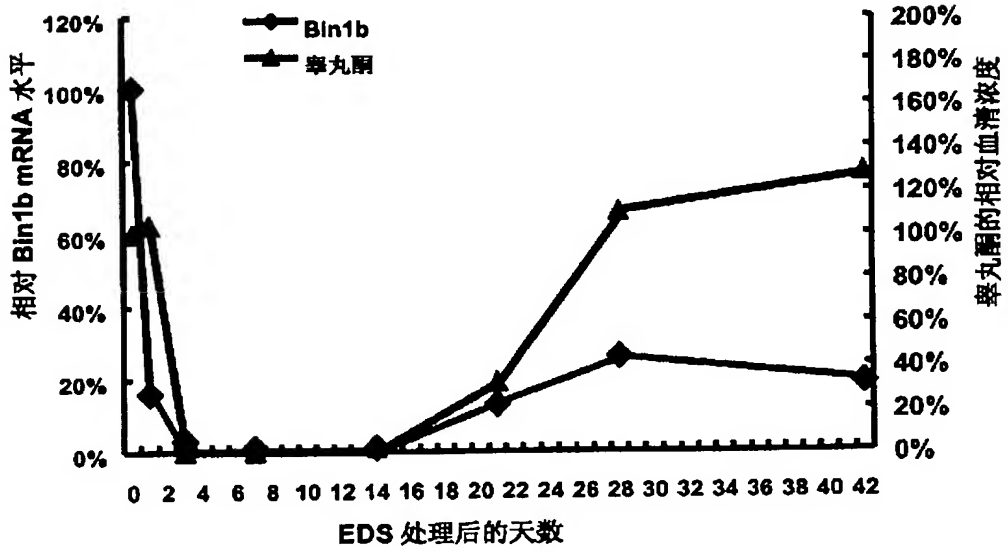

Bin1b
睾丸酮
相对 Bin1b mRNA 水平
睾丸酮的相对血清浓度
120%
100%
80%
60%
40%
20%
0%
200%
180%
160%
140%
120%
100%
80%
60%
40%
20%
0%
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42
EDS 处理后的天数

图 6

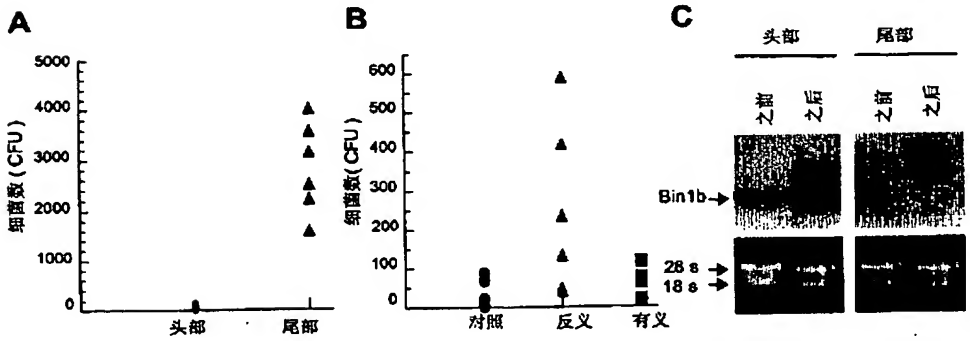

A
B
C
细菌数(CFU)
5000
4000
3000
2000
1000
0
头部
尾部
600
500
400
300
200
100
0
对照
反义
有义
头部
尾部
之前
之后
之前
之后
Bin1b
28 s
18 s

图 7

VERIFICATION OF TRANSLATION

To Whom it may concern:

I, the undersigned technical translator, do hereby declare that I am conversant with the Chinese and English languages. I certify that, to the best of my knowledge and belief, the attached document is a true, accurate and complete translation of the original application document of Chinese Patent Application No. 01105283.X, filed on January 22, 2001 with the Chinese Patent and Trademark Office.

Signature: Xu Xun Date: Oct. 12, 2004

Name: Xu, Xun
Translator

A NOVEL NATURAL ANTIBACTERIAL PEPTIDE, THE NUCLEOTIDE SEQUENCE ENCODING IT AND THE USE THEREOF

Field of invention

This invention relates to molecular biology and immunology, reproductive biology and medicine. In particular, it relates to a novel natural antibacterial peptide specifically expressed in rat caput epididymidis - Bin1b protein, its encoding polynucleotide, and the preparation and uses thereof. Bin1b protein relates to male fertility, especially sperm maturation.

Background

The sequencing of human genome describes code sequences. Which region is a gene? What is the function? Which genes are needed for a biological function? How do they cooperate in normal life? Which mistakes are the causes for diseases? This difficult, challenging, beneficial, and profitable decoding works bring us into an era of functional genomics.

At the turn of the new century, the scientists deem that although the achievements of reproductive biology are great, the prospect of development is pessimistic if only the classic methods limited to tissues and cells are used without utilizing new technologies, without studying the molecular mechanisms, and without combining different disciplines. People have focused on female reproduction and paid little attention to male reproduction. The basic researches on reproductive principles are neglected, resulting in no ideal medicine and techniques for birth control and insufficient knowledge of human reproduction. The reproductive biology encounters two revolutionary changes in the 21st century. One is to strengthen the basic researches and study the mechanism for human reproduction and birth defects by using the new methods of molecular biology, molecular immunology and cell biology. The other is to strengthen the study of male reproductive biology, thereby developing male contraceptives.

China has 22% population, only 7% cultivated land and 6% fresh water in the world. Even the growth rate is 0.01%, 13 million peoples increase each year. The contraception methods, which are imperfect and have some side effects, are mainly used by females. The safe sterilization means are also taken by females. According to WHO statistics, only 1/12 males in Sichuan province are vasoligated. Actually, both men and women have responsibility of contraception. The control of male reproduction is important to human quantity and quality because (1) a healthy man produces 10^8 sperms/day in 50 years and a woman produce 1 egg/month in 40 years, (2) the sperms are sensitive to environment, (3) the quantity and quality of sperm have decreased 40% in the last 50 years, and (4) 5-10% of males under 45 have primary sterility. To correct the bias, it is necessary to strengthen the researches and develop male contraception means. It is well-known that the health of human reproduction is important worldwide. The contraception is only a temporary measure for population crisis, and the off-springs should seek a physically and mentally healthy life. The design of birth-control medicine should suit the new needs. On September 9-10, 1999, in the seminar of "male contraception in 21st century" held by NICHD in NIH headquarters, people studied the situation, put forward the objects and arranged the activities including co-operation and fund support [Trends in Endocrinology and Metabolism 2000, 11 (2):66-69].

Spermatocytes undergo mitosis, meiosis and differentiation and form sperms in testis. Then they enter into epididymis, gradually maturate in caput and corpus, and deposit at cauda until ejaculation. A series of maturation changes including the motor capacity of sperm, the formation of acrosome

function, the conversion of metabolism are not accomplished by sperm themselves. The sperms gradually maturate by interacting with the epididymis micro-environment when passing the epididymis. Epididymis is a long, narrow and zigzag passage connecting testis and spermaductus. The cells in different parts of epididymis express different genes and products, and excrete different proteins and molecules. The different fluid components, ionic strength and pH form the changing micro-environment and interact with sperm to partially alter or modify the sperm surface protein, such as phosphorylation, esterification, acylation, carboxylation, and glycosylation. The sperms gradually obtain the functions and immunologic defense for maturation, which protect sperms in epididymis until they pass female genital tract toward ovum.

The maturation, storage and protection of epididymis have the following properties and, therefore, it is an ideal target for birth control.

(1)The function is simple. The interfering medicines are unlikely to cause severe side effects.

(2)Epididymis is a final organ for hormones and has no endocrine function. The epididymis medicines normally do not effect hormone secretion.

(3) Before entering epididymis, sperms are completely differentiated and transcription is stopped. The maturation involves protein modification but no DNA replication. The medicines are unlikely to cause DNA mutations and diseases.

(4) The study of epididymis is not emphasized and has promising potential.

The mechanism for forming the micro-environment in parts of epididymis is not determined by one or two genes and protein, but by a group of co-operating products. Little is known about the initiation and process of epididymis gene expression related to sperm maturation. The study helps to disclose the molecular mechanism of sperm maturation, decode genomic codes, establish a foundation for sperm-related infertility, and provide new routes for develop male contraception drugs which blocks sperm maturation.

Since 1970's, The researches on epididymis mainly focused on the following three aspects:

(1) Some results have been obtained on protein level by comparing the proteins in the lumen at different parts of epididymis or the differently maturated membrane proteins of sperm at different parts through dielectrophoresis, or by comparing them in immunoassay using polyclonal antibodies against proteins. However, the developments are dissatisfactory due to low sensitivity of separation or analysis technology. The Dacheux Laboratory (France) has reported the identification of more than 200 proteins from the lumen of epididymis of swine and sheep, indicating the research on epididymis is less difficult than the other organs. However, only 15 epididymis specific cDNAs have been cloned.

(2) The known roles of epididymis in sperm maturation may be used to determine whether the specific function of epididymis is related to some known proteins. For example, it is known that epididymis protects sperms from damage caused by oxygen free-radical. The mRNAs of six antioxidases on different parts of epididymis have been detected. E-GPX and E-SOD have the highest amount of mRNA at caput and corpus, respectively, indicating that the different parts of epididymis need different antioxidases. However, those researches are restricted to the known knowledge and hardly helpful to find new function genes or products. Additionally, there are many immunocytes and immuno-depressive molecules in epididymis, which forms a protecting immunological micro-environment. However, the molecules and cellular mechanism for the formation and regulation are unclear.

(3) The molecular biochemistry technology develops quickly since 1990's and one can find new genes specifically expressed in epididymis on mRNA level by subtraction hybridization, etc. A

German laboratory found 6 new mRNAs in human epididymis by subtraction screening of cDNA library. However, due to the limitations of the human epididymis or other materials, their functions were not intensively studied. Therefore, people study the laboratory animals by different means.

Therefore, there is an urgent need to develop new natural protein related to male reproduction and epididymis.

Summary of Invention

One purpose of the invention is to provide a novel antibacterial peptide -Binlb and its fragments, analogs and derivatives.

Another purpose of the invention is to provide polynucleotides encoding the polypeptides.

Still another purpose of the invention is to provide the preparation and uses of the polypeptides and polynucleotides.

In the 1st aspect, the invention provides an isolated rat Binlb polypeptide, which comprises a polypeptide having the amino acid sequence of SEQ ID NOs: 2 or 3, its conservative variants, its active fragments, and its active derivatives. Preferably, said polypeptide has the amino acid sequence of SEQ ID NOs: 2 or 3.

In the 2nd aspect, it provides an isolated polynucleotide comprising a nucleotide sequence sharing at least 70% identity to the following nucleotide sequences: (a) the polynucleotide encoding Binlb polypeptide of SEQ ID NOs: 2 or 3 ; (b) the polynucleotide complementary to polynucleotide of (a). Preferably, said polynucleotide encodes a polypeptide comprising the amino acid sequence of SEQ ID NOs: 2 or 3. More preferably, said polynucleotide is selected from the group consisting of (a) 57-260 of SEQ ID NO:1; (b) 105-260 of SEQ ID NO:1; and (c) 1-336 of SEQ ID NO:1.

In the 3rd aspect, it provides a vector comprising the above polynucleotide, and a host cell transformed with the vector or polynucleotide.

In the 4th aspect, it provides a method for producing a polypeptide having the activity of Binlb protein, which comprises:

(a) culturing the above transformed host cell under the expression conditions;

(b) isolating the polypeptides having the activity of Binlb protein from the culture.

In the 5th aspect, it provides an antibody specifically binding Binlb protein. Also provided are nucleic acid molecules comprising consecutive 15-1757 nucleotides of the above polynucleotide.

In the 6th aspect, it provides compounds that simulate, promote and antagonize Binlb activity, or inhibit Binlb expression and methods for screening and preparing these compounds. Preferably, the compounds are antisense sequences of Binlb encoding sequence or fragments thereof.

In the 7th aspect, it provides a method for detecting Binlb protein in a sample, comprising: contacting the sample with the antibody specifically against Binlb protein, observing the formation of antibody complex which indicates the presence of Binlb protein in the sample.

In the 8th aspect, it provides a method for determining the diseases related to Binlb abnormal expression or the susceptibility thereof, which comprises detecting the mutation of Binlb encoding sequence.

In the 9th aspect, it provides the uses of Binlb and its encoding sequence, e.g., in screening Binlb agonists and antagonist, and peptide fingerprinting. The Binlb encoding sequence and its fragment can be used as primers in PCR, or probes in hybridization and microarray.

In the 10th aspect, it provides a pharmaceutical composition comprising a safe and efficient

amount of Bin1b protein, or its agonist or antagonist and pharmaceutically acceptable carrier. This pharmaceutical composition can be used to treat diseases, e.g., urogenital infection.

In the 11th aspect, it provides a microbicide comprising an antimicrobially efficient amount of Bin1b polypeptide.

The other aspects of invention will be apparent to artisan in light of the teaching of the invention.

The inventors screened the genes specifically expressed in parts of rat and monkey epididymis by differential display and subtraction hybridization and cooperated with North Carolina University on monkey research. We not only identified the known specific genes, but also obtain full-length cDNA clones of genes specifically expressed, 2 in rat caput epididymidis, 4 in monkey epididymis caput, 4 in monkey epididymis corpus, and 3 in monkey epididymis cauda.

Bin1b, which specifically expresses in rat caput epididymidis, is one of them. The full-length cDNA and genomic DNA clones were obtained. The nucleotide and amino acid sequence were registered in Genbank of NIH with accession Nos. AF217088 and AF217089, which will be published after filing this application. Bin1b gene is expressed very specifically, only in the epithelial cells of rat caput epididymidis. It is maximally expressed in sexually mature rats and decreased in old rats, indicating Bin1b relates to reproduction. Androgens up-regulate the Bin1b expression. One may influence Bin1b expression with hormones and design male contraceptives to regulate sperm maturation. Further, Bin1b is the first natural antibacterial peptide in beta-defensin family found in rat epididymis. It has prospect of being developing into a natural drug for curing urogenital infection.

Description of Drawings

The following drawings illustrate the embodiments, and do not limit the scope of invention defined in the claims.

Fig. 1. DD-RT-PCR of rat epididymis caput (1), corpus (2) and cauda (3). The sample are duplicated for accuracy. Bin1b is differentially expressed in caput region as arrow indicated.

Fig. 2A. Sequence and structural characteristics of Bin1b. Genomic DNA sequence (Genebank accession number: AF217089) of Bin1b. It is cloned by PCR using primers located at both ends of the Bin1b full-length cDNA (AF217088) (upper primer 5'-GGACACCCAGTCATCAGTCACAT-3' (SEQ ID NO:9) and lower primer 5'-TTTGGGGTGCTT CCAGGTCTCT-3' (SEQ ID NO:10)) with rat genomic DNA as template. Two exons: uppercase; Coding region: shadowed portion; PolyA signal: bold uppercase; Intron: lowercase; Splicing site: bold lowercase; Putative signal peptide: amino acids underlined. The potential N-terminal is a myristoylated G residue in box whose consensus pattern (GIRNTV) is in bold italic uppercase and S residue in circle may be phosphorylated by PKC whose consensus pattern (SIK) is in bold italic uppercase. Stop codon is designated by *.

Fig. 2B. Sequence Similarity of Bin1b with (-defensins. The conserved six Cysteine residues in (-defensins are shadowed. BNBD9 (AAB25872), BNBD3 (AAB25866), BNBD7 (AAB25870), TAP (P25068), LAP (Q28880), EBD (O02775) from cattle; HBD1 (Q09753), HBD2 (O15263), HBD3 (NP061131) from human; EP2E (AF263555_1) CBD1 (AF188607_1), CBD2 (AF209855_1) from Chimpanzee; MBD1 (AAB72003), MBD2 (CAB42815), MBD3 (AF092929_1), MBD4 (AF155882_1) from mouse; RBD1 (AAC28071), RBD2 (AAC28072) from rat; GBD1 (CAA76811), GBD2 (CAA08905) from goat; SBD1 (O19038), SBD2 (O19039) from sheep; PBD1 (O62697) from pig; Gal1 (P46156), Gal1a (P46157) and Gal2 (P46158) are chicken

gallinacin 1, 1 (and 2 respectively; THP1 (P80391) and THP2 (P80392) are turkey heterophil peptide 1 and 2 respectively. EP2E is one of the chimpanzee homologs of Bin1b, which was not considered as a member of (-defensin family by its original authors.

Fig. 2C. Alignment of Bin1b with its primate homologs. Compared to Bin1b, its primate homolog EP2D (AF263554_1) and HE2 (1 (AF168617_1) have an extended N-terminal and C-terminal and EP2E (AF262555_1) has only extended C-terminal. HE2 (1 is designated as HE2b1 in the figure.

Fig. 3. In vitro transcription and translation assay of Bin1b full-length cDNA.

3A. In vitro transcription and translation assay of Bin1b full-length cDNA. *Lanes 1* and *2* are duplicate samples.

3B. Negative control (no plasmid DNA) and luciferase SP6 and T7 control are assayed in the same system as 3A. but run in 12% SDS-PAGE gel because of molecular weight differences. *Lane 1.* Negative control (no plasmid DNA). *Lane2.* Luciferase SP6 control. *Lane 3.* Luciferase T7 control.

3C. Construction of pSPT18-Bin1b.

Fig. 4. Recombinant Fusion Expression of Bin1b. The fusion protein DHFR-Bin1b is 31 KD as indicated by the arrow. NI: non-induced, I: induced, CL: clear lysate, FT: flow through, W: wash buffer, E1: elution 1, E2: elution 2, M: marker.

Fig. 5. Localization and developmental regulation of Bin1b. **(A)** Tissue distribution of Bin1b by Northern analysis. **(B)** Regional distribution of Bin1b in the epididymis by in situ hybridization. Bin1b is located in middle (Mid) of the rat epididymis caput region. Ins: initial segment. **(C)** Cellular localization of Bin1b using antisense probe. Bin1b is located in the principle cells (Prc) of the epididymis. **(D)** Sense probe. **(E)** 18s probe as positive control. **(F)** The expression profile of Bin-1b mRNA in the rat epididymis during the whole life span (15-720 days). (1, 2 and 3 represent the mRNA from the caput, corpus and cauda regions, respectively)

Fig. 6. Northern blot analysis of rat epididymis caput total RNA after Ethylene Dinethamesulfonate (EDS) treatment.

Fig. 6A. Northern blot analysis picture.

Fig. 6B shows Bin1b expression is up-regulated by androgen. The "▲" line represents rat serum testosterone level at various time intervals after EDS injection. The "◆" line represents rat Bin1b expression changes which are calibrated by 18s level.

Fig. 7. Antimicrobial activity of Bin1b and its expression upregulation in response to inflammation. **(A)** Comparison of antimicrobial activity in caput and cauda cultures. 100 colony forming units (CFU) of *E. coli* was added to cultures 16 hours prior to examination. **(B)** Effect of antisense of Bin1b on antimicrobial activity of the caput cultures. The cultures were transfected with antisense or sense oligos (5 ug/ (l) for 20 hours. **(C)** Enhanced Bin1b mRNA in the caput but not the cauda region of the rat epididymis inflamed by two-week ligation of the spermaductus.

Detailed description

As used herein, the term "Bin1b protein", "Bin1b polypeptide" or "antimicrobial peptide Bin1b" are exchangeable, referring to a protein or polypeptide comprising the amino acid sequence of natural antimicrobial peptide Bin1b (SEQ ID NOs: 2 or 3). The term includes Bin1b with or without the starting Met residue, Bin1b with or without signal peptide. The mature Bin1b is shown in SEQ ID NO:3.

As used herein, the term "isolated" refers to a substance which has been isolated from the

original environment. For naturally occurring substance, the original environment is the natural environment. E.g., the polynucleotide and polypeptide in a naturally occurring state in the viable cells are not isolated or purified. However, if the same polynucleotide and polypeptide have been isolated from other components naturally accompanying them, they are isolated or purified. "Isolation" and "purification" include separating recombinant Binlb protein from other proteins, saccharide, etc.

As used herein, the terms "isolated Binlb protein or polypeptide" mean that Binlb polypeptide does not essentially contain other proteins, lipids, carbohydrate or any other substances associated therewith in nature. The artisans can purify Binlb protein by standard protein purification techniques.

The polypeptide of invention may be a recombinant, natural, or synthetic polypeptide, preferably a recombinant polypeptide. The polypeptide of invention may be a purified natural product or a chemically synthetic product. Alternatively, it may be produced from prokaryotic or eukaryotic hosts, such as bacteria, yeast, higher plant, insect, and mammalian cells, using recombinant techniques. According to the host used in the recombinant production, the polypeptide may be glycosylated or non-glycosylated. The polypeptide may or may not comprise the starting Met residue.

The invention further comprises the fragments, derivatives and analogues of Binlb. As used in the invention, the terms " fragment ", " derivative " and " analogue " mean the polypeptide that essentially retains the same biological functions or activity of natural Binlb protein. The fragment, derivative or analogue of the polypeptide may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue), or (ii) one in which one or more of the amino acid residues include a substituent group, or (iii) one in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to mature polypeptide, such as a leader or secretary sequence or a sequence used for purifying polypeptide or proprotein, e.g., a fusion protein formed with IgC fragment. Such fragments, derivatives and analogs are known to the artisans based on the teachings herein.

The particular Binlb analogs are homologous proteins in other mammals, e.g., cow, sheep, rabbit, dog, monkey, human, etc. One can obtain the sequences encoding these homologous proteins by hybridization or amplification based on the disclosed sequence herein, and obtain the proteins using conventional recombinant techniques.

In the present invention, the term "Binlb polypeptide" means a polypeptide having the activity of Binlb protein comprising the amino acid sequence of SEQ ID NOs: 2 or 3. The term also comprises the variants which have the same function of Binlb. These variants include, but are not limited to, deletions, insertions and/or substitutions of several amino acids (typically 1-20, preferably 1-10, more preferably 1-5, most preferably 1-3), and addition of one or more amino acids (typically less than 20, preferably less than 10, more preferably less than 5) at C-terminal and/or N-terminal. E.g., the protein functions are usually unchanged when an amino residue is substituted by a similar or analogous one. Further, the addition of one or several amino acids at C-terminal and/or N-terminal usually does not change the protein function. The term also includes the active fragments and derivatives of Binlb protein.

The variants of polypeptide include homologous sequences, conservative mutants, allelic variants, natural mutants, induced mutants, proteins encoded by DNA which hybridizes to Binlb DNA under high or low stringency conditions as well as the polypeptides retrieved by antisera raised

against Bin1b polypeptide. The present invention also provides other polypeptides, e.g., fusion proteins, which include the Bin1b polypeptide or fragments thereof. Besides substantially full-length polypeptide, the soluble fragments of Bin1b polypeptide are also included. Generally, these fragments comprise at least 15, typically at least 25, preferably at least 35, more preferably at least 40 consecutive amino acids of Bin1b polypeptide.

The invention also provides the analogues of Bin1b polypeptide. Analogues can differ from naturally occurring Bin1b polypeptide by amino acid sequence differences or by modifications which do not affect the sequence, or by both. These polypeptides include genetic variants, both natural and induced. Induced variants can be made by various techniques, e.g., by random mutagenesis using irradiation or exposure to mutagens, or by site-directed mutagenesis or other known molecular biologic techniques. Also included are analogues which include residues other than those naturally occurring L-amino acids (e.g., D-amino acids) or non-naturally occurring or synthetic amino acids (e.g., beta- or gamma-amino acids). It is understood that the polypeptides of the invention are not limited to the representative polypeptides listed hereinabove.

Modifications (which do not normally alter primary sequence) include in vivo or *in vitro* chemical derivation of polypeptides, e.g., acelylation, or carboxylation. Also included are modifications of glycosylation, e.g., those made by modifying the glycosylation patterns of a polypeptide during its synthesis and processing or in the further processing steps, e.g., by exposing the polypeptide to glycosylation enzymes (e.g., mammalian glycosylating or deglycosylating enzymes). Also included are sequences having phosphorylated amino acid residues, e.g., phosphotyrosine, phosphoserine, phosphothronine, as well as sequences modified to improve the resistance to proteolytic degradation or to optimize solubility properties.

In the invention, "Bin1b conservative mutant" means a polypeptide formed by substituting at most 10, preferably at most 8, more preferably 5, and most preferably at most 3 amino acids with the amino acids having substantially the same or similar property, as compared with the amino acid sequence of SEQ ID NOs: 2 or 3. Preferably, these conservative mutants are formed by the substitution according to Table 1.

Table 1

Initial residue	Representative substitution	Preferred substitution
Ala (A)	Val; Leu; Ile	Val
Arg (R)	Lys; Gln; Asn	Lys
Asn (N)	Gln; His; Lys; Arg	Gln
Asp (D)	Glu	Glu
Cys (C)	Ser	Ser
Gln (Q)	Asn	Asn
Glu (E)	Asp	Asp
Gly (G)	Pro; Ala	Ala
His (H)	Asn; Gln; Lys; Arg	Arg
Ile (I)	Leu; Val; Met; Ala; Phe	Leu
Leu (L)	Ile; Val; Met; Ala; Phe	Ile
Lys (K)	Arg; Gln; Asn	Arg
Met (M)	Leu; Phe; Ile	Leu
Phe (F)	Leu; Val; Ile; Ala; Tyr	Leu

Pro (P)	Ala	Ala
Ser (S)	Thr	Thr
Thr (T)	Ser	Ser
Trp (W)	Tyr; Phe	Tyr
Tyr (Y)	Trp; Phe; Thr; Ser	Phe
Val (V)	Ile; Leu; Met; Phe; Ala	Leu

The polynucleotide of invention may be in the forms of DNA and RNA. DNA includes cDNA, genomic DNA, and synthetic DNA, etc., in single strand or double strand form. A single strand DNA may be an encoding strand or non-encoding strand. The coding sequence for mature polypeptide may be identical to the coding sequence shown in SEQ ID NO:1, or is a degenerate sequence. As used herein, the term "degenerate sequence" means an sequence which encodes a protein comprising the sequence of SEQ ID NOs: 2 or 3 and which has a nucleotide sequence different from the coding region in SEQ ID NO:1.

The sequences encoding the mature polypeptide include those encoding only the mature polypeptide, those encoding mature polypeptide plus various additional encoding sequence, the encoding sequence for mature polypeptide plus the non-encoding sequence and optional additional encoding sequence.

The term "polynucleotide encoding the polypeptide " includes the polynucleotide encoding said polypeptide and the polynucleotide comprising additional and/or non-encoding sequence.

The invention further relates to the variants of polynucleotides which encode a polypeptide having the same amino acid sequence, or its fragment, analogue and derivative. The variant of the polynucleotide may be a naturally occurring allelic variant or a non-naturally occurring variant. Such nucleotide variants include substitution, deletion, and insertion variants. As known in the art, the allelic variant is a substitution form of polynucleotide, which may be a substitution, deletion, and insertion of one or more nucleotides without substantially changing the functions of the encoded polypeptide.

The invention further relates to polynucleotides, which hybridize to the hereinabove-described sequences, if there is at least 50%, preferably at least 70%, and more preferably at least 80% between the sequences. The invention particularly relates to polynucleotides, which hybridize under stringent conditions to the polynucleotides of the invention. As herein used, the term "stringent conditions" means the following conditions: (1) hybridization and washing under low ionic strength and high temperature, such as 0.2xSSC, 0.1% SDS, 60°C; (2) hybridization after adding denaturants, such as 50% (v/v) formamide, 0.1% bovine serum/0.1% Ficoll, 42°C; or (3) hybridization of two sequences sharing at least 95%, preferably 97% homology. Further, the hybridizing polynucleotides encode a polypeptide which retains the same biological function or activity as the mature polypeptide of SEQ ID NO:2

The invention also relates to nucleic acid fragments hybridized with the hereinabove sequence. As used herein, the length of "nucleic acid fragment " is at least 15bp, preferably 30bp, more preferably 50bp, and most preferably at least 100bp. These fragments can be used in the amplification techniques of nucleic acid, e.g., PCR, to determine and/or isolate the Bin1b encoding polynucleotide.

The full-length Bin1b nucleotide sequence or its fragment can be prepared by PCR amplification, recombinant method and synthetic method. For PCR amplification, one can obtain said

sequences by designing primers based on the nucleotide sequence disclosed herein, especially the ORF, and using cDNA library commercially available or prepared by routine techniques in the art as a template. When the sequence is long, it is usually necessary to perform two or more PCR amplifications and link the amplified fragments together correctly.

Once the sequence is obtained, one can produce lots of the sequences by recombinant methods. Usually, said sequence is cloned into a vector which is then transformed into a host cell. The sequence is isolated from the amplified host cells using conventional techniques.

Further, the sequence can be synthesized because Binlb sequence is short. Typically, several small fragments are synthesized and linked together to obtain a long sequence.

It is completely feasible to chemically synthesize the DNA sequence encoding the protein of invention, or the fragments or derivatives thereof. In addition, the mutation can be introduced into the protein sequence by chemical synthesis.

The amplification of DNA/RNA by PCR (Saiki, et al. Science 1985; 230:1350-1354) is preferably used to obtain Binlb gene. Especially when it is difficult to obtain the full-length cDNA, RACE is preferably used. The primers used in PCR can be properly selected according to the sequence information disclosed herein and synthesized by the conventional methods. The amplified DNA/RNA fragments can be isolated and purified by conventional methods, e.g., gel electrophoresis.

The invention further relates to a vector comprising the polynucleotide of invention, a genetic engineered host cell transformed with the vector or the sequence encoding Binlb protein, and the method for producing the Binlb polypeptide by recombinant techniques.

The recombinant Binlb polypeptides can be expressed or produced by the conventional recombinant DNA technology (Science, 1984; 224:1431), using the polynucleotide sequence of invention. Generally, it comprises the following steps:

(1) transfecting or transforming the appropriate host cells with the polynucleotide encoding Binlb polypeptide or the vector containing the polynucleotide ;

(2) culturing the host cells in an appropriate medium;

(3) isolating or purifying the protein from the medium or cells.

In the invention, the polynucleotide sequences encoding Binlb may be inserted into a recombinant expression vector. The term "expression vector" means a bacterial plasmid, bacteriophage, yeast plasmid, plant virus or mammalian cell virus, such as adenovirus, retrovirus or any other vehicles known in the art. Any plasmid or vector can be used to construct the recombinant expression vector as long as it can replicate and is stable in the host. One important feature of expression vector is that the expression vector typically contains a replication origin, a promoter, a marker gene as well as the translation regulatory components.

The known methods can be used to construct an expression vector containing Binlb DNA sequence and appropriate transcription/translation regulatory components. These methods include *in vitro* recombinant DNA technique, DNA synthesis technique, *in vivo* recombinant technique, etc. The DNA sequence is efficiently linked to the proper promoter in an expression vector to direct the synthesis of mRNA. The exemplary promoters are lac or trp promoter of E. coli; P_L promoter of λ phage; eukaryotic promoter including CMV immediate early promoter, HSV thymidine kinase promoter, early and late SV40 promoter, LTRs of retrovirus and some other known promoters which control the gene expression in the prokaryotic cells, eukaryotic cells or virus. The expression vector may further comprise a ribosome-binding site for initiating the translation, transcription terminator and the like.

The expression vector preferably comprises one or more selective marker genes to provide a phenotype for selecting the transformed host cells, e.g., the dehydrofolate reductase, neomycin resistance gene and GFP (green flurencent protein) for eukaryotic cells, as well as tetracycline or ampicillin resistance gene for *E. coli*.

The vector containing said DNA sequence and proper promoter or regulatory elements can be transformed into appropriate host cells to express the protein.

The "host cell " includes prokaryote, e.g., bacteria; primary eukaryote, e.g., yeast; advanced eukaryotic, e.g., mammalian cells. The representative examples are bacterial cells, e.g., *E. coli*, *Streptomyces*, *Salmonella typhimurium*; fungal cells, e.g., yeast; plant cells; insect cells e.g., Drosophila S2 or Sf9; animal cells e.g., CHO, COS or Bowes melanoma, etc.

Transcription of the polynucleotide in higher eukaryotes is increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about 10-300bps, that act on a promoter to increase gene transcription. Examples include SV40 enhancer on the late side of replication origin 100 to 270 bp, the polyoma enhancer on the late side of replication origin, and adenovirus enhancers.

The artisans know clearly how to select appropriate vectors, promoters, enhancers and host cells.

Recombinant transformation of host cell with the DNA might be carried out by conventional techniques known to the artisans. Where the host is prokaryotic, e.g., E. coli, the competent cells capable of DNA uptake, can be prepared from cells harvested after exponential growth phase and subsequently treated by the $CaCl_2$ method using known procedures. Alternatively, $MgCl_2$ can be used. The transformation can also be carried out by electroporation. When the host is an eukaryote, transfection of DNA such as calcium phosphate co-precipitates, conventional mechanical procedures e.g., micro-injection, electroporation, or liposome-mediated transfection may be used.

The transformants are cultured conventionally to express Binlb polypeptide. According to the used host cells, the medium for cultivation can be selected from various conventional mediums. The host cells are cultured under a condition suitable for its growth until the host cells grow to an appropriate cell density. Then, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period.

In the above methods, the recombinant polypeptide may be included in the cells, or expressed on the cell membrane, or secreted out. If desired, the physical, chemical and other properties can be utilized in various isolation methods to isolate and purify the recombinant protein. These methods are well-known to the artisans and include, but are not limited to conventional renaturation treatment, treatment by protein precipitant (e.g., salt precipitation), centrifugation, cell lysis by osmosis, sonication, supercentrifugation, molecular sieve chromatography or gel chromatography, adsorption chromatography, ion exchange chromatagraphy, HPLC, and any other liquid chromatagraphy, and the combination thereof.

The recombinant Binlb polypeptide have various uses including, but not limited to: curing urogenital infection, and screening out antibodies, polypeptides or ligands as agonists or antagonists of Binlb. The expressed Binlb protein can be used to screen polypeptide library to find out therapeutically valuable polypeptide molecules which inhibit or activate Binlb protein.

In another aspect, the invention also includes polyclonal and monoclonal antibodies (mAbs), preferably mAbs, which are specific for polypeptides encoded by Binlb DNA or fragments thereof. By "specificity", it means an antibody which binds to the Binlb gene products or a fragments thereof. Preferably, the antibody binds to the Binlb gene products or fragments thereof and does not

substantially recognize nor bind to other antigenically unrelated molecules. Antibodies which bind to Bin1b and block Bin1b protein and those which do not affect the Bin1b function are included in the invention.

The invention includes intact monoclonal or polyclonal antibodies, and immunologically-active antibody fragments, e.g., a Fab' or $(Fab)_2$ fragment, an antibody heavy chain, an antibody light chain,, or a chimeric antibody.

The antibodies in the present invention can be prepared by various techniques known in the art. E.g.,, purified Bin1b gene products, or its antigenic fragments can be administrated to animals (e.g., rabbit, mice and rat) to produce polyclonal antibodies. Similarly, cells expressing Bin1b or its antigenic fragments can be used to immunize animals to produce antibodies. Various adjuvants, e.g., Freund's adjuvant, can be used to enhance immunization.

The mAbs can be prepared using hybridoma technique (Kohler et al.,, Nature 256;495, 1975; Kohler et al.,, Eur.J.Immunol. 6:511, 1976; Kohler et al.,, Eur.J.Immunol. 6:292, 1976; Hammerling et al.,, In Monoclonal Antibodies and T Cell Hybridomas, Elsevier, N.Y., 1981). Antibodies comprise those which block Bin1b function and those which do not affect Bin1b function. Antibodies can be produced by routine immunology techniques and using fragments or functional regions of Bin1b gene product prepared by recombinant methods or synthesized by a polypeptide synthesizer. The antibodies binding to unmodified Bin1b gene product can be produced by immunizing animals with gene products produced by prokaryotic cells (e.g., E. coli), and the antibodies binding to post-translationally modified forms thereof can be acquired by immunizing animals with gene products produced by eukaryotic cells (e.g., yeast or insect cells).

The antibody against Bin1b can be used in immunohistochemical method to detect the presence of Bin1b protein in biopsy specimen. The mAb can be radiolabelled and injected into body to trace the position and distribution of Bin1b.

The substances which act with Bin1b protein, e.g., receptors, inhibitors, agonists and antagonists, can be screened out by various conventional techniques, using Bin1b protein.

The Bin1b protein, antibody, inhibitor, agonist or antagonist of the invention provide different effects when administrated in therapy. Usually, these substances are formulated with a non-toxic, inert and pharmaceutically acceptable aqueous carrier. The pH typically is about 5-8, preferably 6-8, although pH may alter according to the property of the formulated substances and the diseases to be treated. The formulated pharmaceutical composition is administrated in conventional routes including, but not limited to, intramuscular, intravenous, subcutaneous, intradermal or topical administration.

The Bin1b polypeptide can be directly used for curing disorders, e.g., urogenital infection. The Bin1b protein can be administrated in combination with other medicaments, e.g, antibiotics including penicillin.

The invention also provides a pharmaceutical composition comprising safe and effective amount of Bin1b protein in combination with a pharmaceutically acceptable carrier. Such a carrier includes but is not limited to saline, buffer solution, glucose, water, glycerin, ethanol, or the combination thereof. The pharmaceutical formulation should be suitable for delivery method. The pharmaceutical composition may be in the form of injections which are made by conventional methods, using physiological saline or other aqueous solution containing glucose or auxiliary substances. The pharmaceutical compositions in the form of tablet or capsule may be prepared by routine methods. The pharmaceutical compositions, e.g., injections, solutions, tablets, and capsules, should be manufactured under sterile conditions. The active ingredient is administrated in

therapeutically effective amount, e.g., about 1ug - 5mg/kg body weight per day. Moreover, the polypeptide of invention can be administrated together with other therapeutic agents.

When using pharmaceutical composition, the safe and effective amount of the Binlb protein or its antagonist or agonist is administrated to mammals. Typically, the safe and effective amount is at least about 1 ug/kg body weight and less than about 8 mg/kg body weight in most cases, and preferably about 10ug-1mg/kg body weight. Certainly, the precise amount depends upon various factors, such as delivery methods, the subject health, etc., and is within the judgment of the skilled clinician.

The polypeptide molecule capable of binding Binlb protein can be obtained by screening out the random polypeptide library consisting of the various combinations of amino acids bound onto the solid matrix.

The invention further provides diagnostic assays for quantitative and *in situ* measurement of Binlb protein level. These assays are known in the art and include FISH assay and radioimmunoassay. The level of Binlb protein detected in the assay can be used to illustrate the importance of Binlb protein in diseases and to determine the Binlb-related diseases.

A method of detecting Binlb protein in a sample by utilizing the antibody specifically against Binlb protein comprises the steps of: contacting the sample with the antibody specifically against Binlb protein; observing the formation of antibody complex which indicates the presence of Binlb protein in the sample.

The polynucleotide encoding Binlb protein can be used in the diagnosis and treatment of Binlb related diseases. In diagnosis, the polynucleotide encoding Binlb can be used to detect whether Binlb is expressed or not, and whether the expression is normal or abnormal in the case of diseases. Binlb DNA sequences can be used in the hybridization with biopsy samples to determine Binlb expression. The hybridization methods include Southern blotting, Northern blotting and *in situ* blotting, etc., which are public and sophisticated techniques. The corresponding kits are commercially available. A part of or all of the polynucleotides of the invention can be used as probe and fixed on a microarray or DNA chip for analyzing the differential expression of genes in tissues and for the diagnosis of genes. The Binlb specific primers can be used in RT-PCR and *in vitro* amplification to detect the transcripts of Binlb.

Detection of Binlb gene mutation is useful for the diagnosis of Binlb related diseases. The mutation forms of Binlb include site mutation, translocation, deletion, rearrangement and any other mutations compared with the wild-type Binlb DNA sequence. The conventional methods, e.g., Southern blotting, DNA sequencing, PCR and *in situ* blotting, can be used to detect mutation. Moreover, mutation sometimes affects protein expression. Therefore, Northern blotting and Western blotting can be used to indirectly determine the gene mutation.

The sequences of invention are also valuable for chromosome identification. Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-35 bp) from the Binlb cDNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only the cell hybrids, which contain the genes corresponding to the primers, produce amplified fragments.

Once a sequence is mapped to a precise chromosomal location, the physical position of the sequence in chromosome can be correlated with genetic map data. Such data are found in, e.g., Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationships between genes and diseases that have been mapped to the same

chromosomal region are then identified through linkage analysis.

The Bin1b provides approach for curing urogenital system diseases and developing male contraception, thus having huge potential applications.

The invention is further illustrated by the following examples. These examples are only intended to illustrate the invention, but not to limit the scope of the invention. For the experimental methods in the following examples, they are performed under routine conditions, e.g., those described by Sambrook. et al., in Molecule Clone: A Laboratory Manual, New York: Cold Spring Harbor Laboratory Press, 1989, or as instructed by the manufacturers, unless otherwise specified.

Example 1

Discovery of Bin1b differential cDNA fragment

Mainly following the procedures described by Liang, P. et al.[P. Liang, and A.B. Pardee, Science 257, 967 (1992)], total RNAs were isolated from the caput, corpus and cauda regions of the adult Sprague-Dawley rat epididymis, using RNase-free DNase digestion to remove residual chromosome DNA. After precipitation, retrotranscription was carried out with 2 ug isolated total DNAs and 2.5 uM lower primer $T_{11}CA$ with 400 units of MMLV reverse transcriptase (Gibco, BRL) were used. 1/20 of the reverse transcription product was used as template to perform the PCR with 2.5 uM of the upper primer 502 (5'-TGGATTGGTC-3') and 0.5 uM of the lower primer $T_{11}CA$ in a 20 ul volume containing 10 mM Tris-HCl pH 9.0, 1.5 mM $MgCl_2$, 50 mM KCl, 0.1% Triton X-100, 4 uM each dNTP, 1 uCi ^{32}P-dATP and 3 units of Taq polymerase. The PCR conditions were 94 ^{0}C for 5 min, then 40 cycles of 94 ^{0}C for 30 sec, 40 ^{0}C for 60 sec, and 72 ^{0}C for 50 sec, with a final elongation at 72 ^{0}C for 10 min. PCR products were precipitated and fractionated on the 0.2 mm thick 6% sequencing gel.

The band (~340 bp) differentially displayed only in the RNA from caput region was named Bin1b (Fig. 1)and cut out and left in boiling distilled water for 15min. The eluted DNAs were precipitated with ethanol and re-amplified in the same conditions as described above except that 40 uM instead of 4 uM each dNTP was used and no radioisotope was involved. The resultant 340 bp DNA fragments were cloned into pBluescript SK (plasmid. 52 clones were screened by reverse Northern and the clone having specific hybridization signal in the caput region was named Bin-1b. The full-length cDNA of Bin1b was cloned by 5'-RACE based on said fragment.

Example 2

Cloning and characterization of full-length Bin1b cDNA:

Primers were designed based on obtained cDNA fragment for extending the 5' end. Two 5'-RACE approaches were used to obtain full-length cDNA.

The first approach was a conventional method reported by A. N. Apte and P. D. Siebert [in Reverse Transcriptase PCR, J. W. Larrick, P.D. Siebert, Eds. (Horwood, London, 1995), pp. 232-244], using DNA oligo-first strand cDNA ligation.

The other was with DNA oligo –RNA ligation [modified from K, Maruyama, S. Sugano, Gene 138, 171 (1994), to ensure the obtaining of the 5'-capping site]. Oligo DNA was used instead of oligo RNA to ligate with total RNA. The procedure was as follows:

Total RNA (50 ug) from rat epididymis caput was digested with 400 units of bacterial alkaline phosphatase (BAP) at 37 ^{0}C for 30 min and additional incubation at 65 ^{0}C for 30 min. The BAP was

then digested with Proteinase K (50 ng/ (l) at 37 ^{0}C for 30min. After purification, 10ug of RNA was further treated with 2 units of tobacco acid pyrophosphatase (TAP) at 37 ^{0}C for 2 hours, extracted with phenol/chloroform, and precipitated with ethanol. The untreated RNA, BAP-treated RNA, and TAP-treated RNA (0.75 ug, about 3 pmol, respectively) were mixed with 1.25 pmol of the DNA oligonucleotide 7209 (5'- AATGGTACCGT- GACGTGGTCC-3') (SEQ ID NO:5)and ligated with T4 RNA ligase (1.2 unit/ ul) in 10 ul of 50 mM Tris- HCl (pH 8.0), 10mM $MgCl_2$, 1mM hexamine colbalt chloride, 25% PEG 8000, and 1 mM adenosine triphosphate at 17^{0}C for 18 hours. Superscript II One-Step RT- PCR system (Gibco, BRL) was used to perform the RT-PCR reaction with 0.2 ul of the ligation products in 20 ul containing 200mM oligo DNA #7209 (5'-AATGGTACCGTGACGTGGTCC-3', SEQ ID NO:5) and 200 mM Binlb gene-specific primer (GSP) (5'-TGGCCCCGCTGCATGAAGCAC-3', SEQ ID NO:6). The RT-PCR reactions were as follows: 50 ^{0}C for 30 min, 94 ^{0}C for 2 min, then 35 cycles of 94 ^{0}C for 5s, 60 ^{0}C for 15s, and 72 ^{0}C for 45s, with a final elongation 72 ^{0}C for 5 min. The second PCR was performed with 0.2ul of the RT-PCR products in 10ul volume containing 50 mM Tris- HCl (pH 8.3), 1-3 mM $MgCl_2$, bovine serum albumin (250 ug/ ml), 0.5% Ficoll 400, 1mM tartrazine, 200uM dNTP, 500 nM Binlb GSP 500nM oligo DNA #7209, and 0.4 unit of Taq polymerase. PCR reactions were run in capillary tubes as follows: 94 ^{0}C for 1 min, then 60 cycles of 94 ^{0}C for 0s, 60 ^{0}C for 0s, and 77 ^{0}C for 15s, with a final elongation at 77 ^{0}C for 5 min. The PCR product was cloned into pBluescript SK^+ T-Vector and sequenced. Thus, the 3' and 5' fragments of Binlb were obtained.

Primers was designed on both terminals of cDNA fragments (5'-GGACACCCAG TCATCAGTCA- 3' (SEQ ID NO:7) and 5'-CAGCAAGTGT TTATTGAGCA-3' (SEQ ID NO:8)). Using RT-PCR product as a template, Binlb full-length cDNA clone was obtained by PCR. The sequence was confirmed in 6 rats of different ages. Binlb full-length cDNA was 385bp (SEQ ID NO:1 and Fig. 2A) which encodes a 68aa peptide (named Binlb protein) (SEQ ID NO:2) with a signal peptide of 16 amino acids at N-terminal. The putative mature protein contains 45 amino acid (another 7 amino acid was removed during pro-Binlb maturation).

The result of blast search indicated the coding sequence (180-241) of Binlb had 83% identity with noncoding sequence (456-518) of human sperm antigen HE2 (673bp). The peptide coded by Binlb exhibited some similarity with mammalian beta-defensins. (Fig. 2B)

The Binlb genomic DNA was cloned by similar PCR procedure (SEQ ID NO:4, Fig. 2A). The Binlb genomic DNA contained two exons separated by an intron, which was characteristic in beta-defensin gene family.

Several isomers of human HE2 and its homologs in chimpanzee, EP2 were reported by several laboratories. Though some of them were low expressed, they shared high sequence homology with Binlb (Fig. 2C).

Example 3

Expression of Binlb protein

The size of Binlb peptide was confirmed to be about 31 kD by in vitro transcription and translation, meeting the theoretical prediction.

Plasmid pSPT18-Binlb (Fig. 3C) was constructed by cloning Binlb full-length cDNA (including 75bp polyA tail) encoding ORF of 68 amino acids (7799 Dalton) between EcoRI and HindIII sites (downstream of T7 promoter) of pSPT18 (Promega). The plasmid was completely digested by HindIII. The reaction was performed with TNT T7 Coupled Reticulocyte Lysate System

labeled by ^{35}S-Met and then analyzed by 16.5% Tricine-SDS-PAGE as described by Schagger and Von Jagow. The result was shown in Fig. 3A. The result was compared with negative control without plasmid DNA and luciferase SP6 and T7 (Fig. 3B).

Example 4

Fusion expression of Bin1b and antibody preparation

A fragment of Bin1b cDNA (126-281) was cloned into fusion expression vector pQE-40 (QIAGEN). The positive clone was induced with 1mM IPTG and purified by Ni-NTA agarose (Qiagen) as described by the manufacturer (see QiaExpressionist, Qiagen).

The purified fusion protein DHFR-Bin1b is 31KD (Fig. 4). Rabbit was immunized with the purified protein to produce the antiserum to DHFR-1b. But the titer of the antiserum to Bin1b was still to be improved.

Example 5

Tissue distribution of Bin1b

Tissue distribution analysis of Bin1b was done by Northern blot following the method of Church and Gilbert (1984). Total RNA was extracted from epididymis caput, testis, ventral prostate, liver, heart, spleen, lung, kidney, thymus, adrenal gland, coagulating gland, seminal vesicle, hypothalamus, striatum, pituitary, hippocampus, cerebellum, cerebral cortex and epididymis caput of Sprague-Dawley rat, respectively. 20ug each of total RNA was separated on 1.2% formaldehyde agarose gel electrophoresis and transferred to Hybond-N+ membrane (Amersham Pharmacia Biotech). Bin1b probe was labeled by Prime-a-Gene System kit (Promega). The result of hybridization indicated that Bin1b was only expressed on rat caput epididymis (Fig. 5A). In situ hybridization was used to decide the precise localization of Bin1b in epididymis.

Example 6

Localization of Bin1b mRNA in epididymis

In this example, conventional *in situ* hybridization was used to map Bin1b. Microscopy showed that Bin1b was localized to the epithelial of middle caput epididymis. (Figs. 5B-E)

Example 7

Developmental change of Bin1b during rat growth

Developmental change of Bin1b was analyzed by in the same northern blot as in Example 5. RNA was extracted from caput epididymis of 15 day-old rat; caput, corpus and cauda epididymis of 30, 45, 60, 120, 270 and 720 day-old rat.

The result showed that expression of Bin1b was highest at sexually mature period (including sexually active period) and gradually decrease afterwards (Fig. 5F). This indicated that Bin1b might be associated with sperm maturation and regulate Bin1b for male contraception.

Example 8

Bin1b expression is regulated by small moleculars such as hormones.

In this example, the study on EDS rat model showed that Bin1b expression was partially up-regulated by androgen.

Adult Sprague-Dawley was intraperitoneally injected with EDS (7.5 mg/ 100g bodyweight) in

DMSO- H_2O (1:3, v/v). The rats were sacrificed on day 1, 3, 7, 14, 21, 28 and 42. The rat without EDS injection served as the negative control. 20 ug total RNA was extracted from rat caput epididymis, fractionated by 1.2% formaldehyde agarose gel electrophoresis, blotted onto Hybond-N+ membrane and hybridized with a probe located at Bin1b cDNA 3' terminal and 18s ribosome RNA probe.

The secretion of androgen decreased immediately after EDS injection but restored in about two weeks after injection. The expression of Bin1b was up-regulated by androgen compared with variation of androgen level (Figs. 6A and 6B).

It is possible to control sperm maturation by regulation of Bin1b expression with small hormone molecule, which pave new way for developing male contraceptives.

Example 9

Bin1b is a native antimicrobial peptide

The studies above indicated Bin1b was a member of beta-defensin gene family and involved in sperm maturation. The antimicrobial activity was confirmed in this Example.

The antimicrobial activity of Bin1b was tested with the secretions of primary cultures of caput epididymal epithelial. A total of 100 colony-forming units (CFU) of *E.coli* was added to the apical compartment of the epithelial cells separately from caput epididymis and cauda epididymis. The medium was collected 16 hours later for CFU counting. Strong bactericidal activity was detected in the medium collected from caput culture, but none in cauda culture, which confirmed that secretion from caput epididymis cell had the antimicrobial activity (Fig. 7A).

To confirm that the antimicrobial activity was indeed contributed by Bin1b although it was expressed in the caput epididymis, the inventors designed antisense RNAs of Bin1b and added them into cultures to block Bin1b expression 24 hours before adding *E. coli*. The antibacterial capability of the caput culture was greatly attenuated, proving the antimicrobial activity of Bin1b. (Fig. 7B)

These results indicate that Bin1b was a novel native antimicrobial peptide in beta-defensin family.

Example 10

Bin1b Expression was up-regulated by inflammation

When inflammation resulted from ligation of spermaductus which led to accumulation of sperm in epididymis occurred, Bin1b mRNA in caput epididymis increased to 3 folds of normal level before the ligation. No change was seen in cauda epididymis (Fig. 7C). This indicated that Bin1b expression was up-regulated by inflammation.

All the documents cited herein are incorporated into the invention as reference, as if each of them is individually incorporated. Further, it would be appreciated that, in the above teaching of invention, the skilled in the art could make certain changes or modifications to the invention, and these equivalents would still be within the scope of the invention defined by the appended claims of the application.

SEQUENCE LISTING

(1)General information:
(ii)Title of invention: A NOVEL NATURAL ANTIBACTERIAL PEPTIDE, THE NUCLEOTIDE SEQUENCE ENCODING IT AND THE USE THEREOF
(iii)Number of Sequences: 10

(2) INFORMATION FOR SEQ ID NO: 1:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH: 385bp
(B)TYPE: nucleic acid
(C)STRANDEDNESS: double
(D)TOPOLOGY: linear
(ii)MOLECULAR TYPE: cDNA
(xi)SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGACACCCAG TCATCAGTCA CATCTGCTTT CCTGCACAGA GAGAGCGCCA TAAAACATGA 60
AGGTTTTGTT ACTCTTTGCT GTTTTCTTCT GCTTGGTCCA AAGAAACTCA GGGGACATAC 120
CACCTGGAAT CAGAAACACC GTGTGCTTCA TGCAGCGGGG CCACTGTAGG CTCTTCATGT 180
GCCGTTCTGG GGAGAGAAAG GGGGATATTT GCTCTGACCC CTGGAACAGA TGCTGCGTAT 240
CCAGTTCCAT TAAAAACAGA TGATAGAAGA CTCATTGGAA GATCTGAGAT GTGGGGTGCA 300
AGCTCTTGGA AGCTAGAGAC CTGGAAGCAC CCCAAAGGCT TTGAGTATGT GTGGCTAATG 360
GTGCGTGCTC AATAAACACT TGCTG 385

(2) INFORMATION FOR SEQ ID NO: 2:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 68 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: polypeptide
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

MKVLLLFAVF FCLVQRNSGD IPPGIRNTVC FMQRGHCRLF MCRSGERKGD 50
ICSDPWNRCC VSSSIKNR 68

(2) INFORMATION FOR SEQ ID NO: 3:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: polypeptide
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GIRNTVCFMQ RGHCRLFMCR SGERKGDICS DPWNRCCVSS SIKNR 45

(2) INFORMATION FOR SEQ ID NO: 4:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH: 1696bp
(B)TYPE: nucleic acid
(C)STRANDEDNESS: double
(D)TOPOLOGY: linear
(ii)MOLECULAR TYPE: genomic DNA
(xi)SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GGACACCCAG TCATCAGTCA CATCTGCTTT CCTGCACAGA GAGAGCGCCA TAAAACATGA 60
AGGTTTTGTT ACTCTTTGCT GTTTTCTTCT GCTTGGTCCA AAGAAACTCA Ggtaaatgtc 120
ttctgagtag ccctggagaa ggcaggatgc ccttttaggt ttgtagacca cattgaggtg 180
tgtccaggta tcaacattgg gcacagatgg tgggccactc tggggctcag ggtcggacca 240
ctttcctaac gaagaggttt tattttgatt ttttttgttt tgttcatttg tcaagagttg 300
caaattttac agcacggaga cacagaggcc tatattctcc attgtgaata agaaggtctg 360
attgtaactt gagagtttat tcaggacaga attacagccg tacctgtgtc aaaagtgtaa 420
ttttactgcc tcgctgtgag cagagaaggt gttcacattt atgcccccttc cctacccatt 480

acatccacag aacaccagat gtatgcttta aatgaatttt caaatgagag aaaaataggt 540
tcctttaaga aagctagagt ccaggtcctg aagccttgaa ttgctggcag ttctgtcaag 600
gtggactaca cccacatctc catgaacctt cccaaccatg gtaaaccgga tgaacacagt 660
atcacaaatc agtccccagc tgaagtccgg ctattgcagg agaccagttt cctaaatgtt 720
acaggcatag gttgggccgc tgttgctttt taacacaggg tgtgcaacat tgttaaaaag 780
gtttttttta accatctctt tcccatggtg ctttcttttg ggggactcta gttgtttttg 840
ttttgtttta ttgttttact tagaaggaca cacaagacac attgttatct ttcttcttct 900
tattgtagtc ataagagtga aaacccaacc atgagctgag acagacccgc tcctaacttt 960
tctatggcct gagacccagc tcctgttatt ctgttctgtt ttctttttct tttttaattt 1020
atttatttta tgtatgtgag tacagtgtca ctgtcttcag acacaccaga agagggtgtc 1080
agatcccatt acagatggtt gtgagccacc atgtggttgc tgggaattga actcgggacc 1140
tctggaagag cagtcagtgc tcttaaccgc tgagccatct ttccagcccc tgttctgttt 1200
tcttaaatac cactcccccа ctccacaatg tacctctatc tctgggcagc tgcagagccc 1260
tggcctgcaa tgggctaggt gacttcacac tcagtctgtc atgccatccc cgaaacacca 1320
cgagatataa atggttgcta ttgaaagcta aggaggaaaa tctcagtgac gccgaaactc 1380
tggaagagtg gagcagattc ttcgagaggg gctgggggct gggggctggg ggctggagcc 1440
actgttttat ctcagtctgt tgtttccaca gGGGACATAC CACCTGGAAT CAGAAACACC 1500
GTGTGCTTCA TGCAGCGGGG CCACTGTAGG CTCTTCATGT GCCGTTCTGG GGAGAGAAAG 1560
GGGGATATTT GCTCTGACCC CTGGAACAGA TGCTGCGTAT CCAGTTCCAT TAAAAACAGA 1620
TGATAGAAGA CTCATTGGAA GATCTGAGAT GTGGGGTGCA AGCTCTTGGA AGCTAGAGAC 1680
CTGGAAGCAC CCCAAA 1696

(2) INFORMATION FOR SEQ ID NO: 5:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH: 21bp
(B)TYPE: nucleic acid
(C)STRANDEDNESS: single
(D)TOPOLOGY: linear
(ii)MOLECULAR TYPE: oligonucleotide
(xi)SEQUENCE DESCRIPTION: SEQ ID NO: 5:
AATGGTACCG TGACGTGGTC C 21

(2) INFORMATION FOR SEQ ID NO: 6:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH: 21bp
(B)TYPE: nucleic acid
(C)STRANDEDNESS: single
(D)TOPOLOGY: linear
(ii)MOLECULAR TYPE: oligonucleotide
(xi)SEQUENCE DESCRIPTION: SEQ ID NO: 6:
TGGCCCCGCT GCATGAAGCA C 21

(2) INFORMATION FOR SEQ ID NO: 7:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH: 20bp
(B)TYPE: nucleic acid
(C)STRANDEDNESS: single
(D)TOPOLOGY: linear
(ii)MOLECULAR TYPE: oligonucleotide
(xi)SEQUENCE DESCRIPTION: SEQ ID NO: 7:
GGACACCCAG TCATCAGTCA 20

(2) INFORMATION FOR SEQ ID NO: 8:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH: 20bp
(B)TYPE: nucleic acid
(C)STRANDEDNESS: single
(D)TOPOLOGY: linear

(ii)MOLECULAR TYPE: oligonucleotide
(xi)SEQUENCE DESCRIPTION: SEQ ID NO: 8:
CAGCAAGTGT TTATTGAGCA 20

(2) INFORMATION FOR SEQ ID NO: 9:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH: 23bp
(B)TYPE: nucleic acid
(C)STRANDEDNESS: single
(D)TOPOLOGY: linear
(ii)MOLECULAR TYPE: oligonucleotide
(xi)SEQUENCE DESCRIPTION: SEQ ID NO: 9:
GGACACCCAG TCATCAGTCA CAT 23

(2) INFORMATION FOR SEQ ID NO: 10:
(i)SEQUENCE CHARACTERISTICS:
(A)LENGTH: 22bp
(B)TYPE: nucleic acid
(C)STRANDEDNESS: single
(D)TOPOLOGY: linear
(ii)MOLECULAR TYPE: oligonucleotide
(xi)SEQUENCE DESCRIPTION: SEQ ID NO: 10:
TTTGGGGTGC TTCCAGGTCT CT 22

What is claimed is:

1. An isolated polypeptide comprising the amino acid sequence of SEQ ID NOs: 2 or 3.

2. The polypeptide of Claim 1 wherein the sequence of the polypeptide is SEQ ID NOs: 2 or 3.

3. An isolated polynucleotide comprising a nucleotide sequence sharing at least 70% homology to a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence encoding a polypeptide comprising the amino acid sequence of SEQ ID NOs: 2 or 3;

(b) the polynucleotide complementary to nucleotide sequence of (a).

4. The polynucleotide of Claim 3 which encodes a polypeptide comprising the amino acid sequence of SEQ ID NOs: 2 or 3.

5. A vector containing the polynucleotide of Claim 3.

6. A genetically engineered host cell comprising the vector of Claim 5.

7. A method for producing a polypeptide having the activity of Bin1b protein, which comprises:

(a) culturing the host cell of Claim 6 under the expression conditions;

(b) isolating the polypeptides having the activity of Bin1b protein from the culture.

8. An antibody specifically bound with the Bin1b polypeptide of Claim 1.

9. A pharmaceutical composition comprising a safe and efficient amount of the polypeptide of Claim 1 and a pharmaceutically acceptable carrier.

10. A microbicide comprising an antimicrobially efficient amount of polypeptide of Claim 1.

Abstract

The invention provides a novel Binlb protein, and its encoding polynucleotide. Binlb protein is a natural antimicrobial peptide and associates to sperm maturation. The invention also discloses the preparation and uses of Binlb protein and nucleic acid. Binlb protein is useful to treat various diseases, e.g, urogenital infection. The invention also provides a pharmaceutical composition containing Binlb protein.

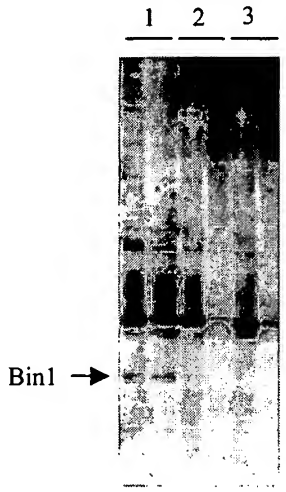


Fig. 1

```
   1  GACACCCAGTCATCAGTCACATCTGCTTTCCTGCACAGAGAGAGCGCCATAAAACATGAAGGTTTTGTTACTCTTTGCTGTTTTCTTCTGCTTGGTCCA
                                                                M  K  V  L  L  L  F  A  V  F  F  C  L  V  Q
 100 AAGAAACTCAG
       R  N  S
                gtaaatgtcttctgagtagccctggagaaggcaggatgccctttaggtttgtagaccacattgaggtgtgtccaggtatcaacattgg
 200 gcacagatggtgggccactctggggctcagggtcggaccactttcctaacgaagaggttttattttgatttttttttgtttgttcatttgtcaagagttg
 300 caaattttacagcacggagacacagaggcctatattctccattgtgaataagaaggtctgattgtaacttgagagtttattcaggacagaattacagccg
 400 tacctgtgtcaaaagtgtaattttactgcctcgctgtgagcagagaaggtgttcacatttatgccccttccctacccattacatccacagaacaccagat
 500 gtatgctttaaatgaattttcaaatgagagaaaaatagqttcctttaagaaagctagagtccaggtcctgaagccttgaattgctggcagttctgtcaag
 600 gtggactacacccacatctccatgaaccttcccaaccatggtaaaccggatgaacacagtatcacaaatcagtccccagctgaagtccggctattgcagg
 700 agaccagtttcctaaatgttacaggcataggttgggccgctgttgctttttaacacagggtgtgcaacattgttaaaaaggttttttttaaccatctctt
 800 tcccatggtgctttctttggggactctagttgttttgttttgttttattgttttacttagaaggacacacaagacacattgttatctttcttcttct
 900 tattgtagtcataagagtgaaaacccaaccatgagctgagacagacccgctcctaacttttctatggcctgagacccagctcctgttattctgttctgtt
1000 ttctttttcttttttaatttatttattttatgtatgtgagtacagtgtcactgtcttcagacacaccagaagagggtgtcagatcccattacagatggtt
1100 gtgagccaccatgtggttgctgggaattgaactcgggacctctggaagagcagtcagtgctcttaaccgctgagccatctttccagcccctgttctgttt
1200 tcttaaataccactcccccactccacaatgtacctctatctctgggcagctgcagagccctggcctgcaatgggctaggtgacttcacactcagtctgtc
1300 atgccatccccgaaacaccacgagatataaatggttgctattgaaagctaaggaggaaaatctcagtgacgccgaaactctggaagagtggagcagattc
1400 ttcgagaggggctgggggctgggggctgggggctggagccactgttttatctcagtctgttgtttccacag
                                                                G  G  D  I  P  P  G  I  R  N  T
1500 GTGTGCTTCATGCAGCGGGGCCACTGTAGGCTCTTCATGTGCCGTTCTGGGGAGAGAAAGGGGGATATTTGCTCTGACCCCTGGAACAGATGCTGCGTAT
       V  C  F  M  Q  R  G  H  C  R  L  F  M  C  R  S  G  E  R  K  G  D  I  C  S  D  P  W  N  R  C  C  V
1600 CCAGTTCCATTAAAAACAGATGATAGAAGACTCATTGGAAGATCTGAGATGTGGGGTGCAAGCTCTTGGAAGCTAGAGACCTGGAAGCACCCCAAAGGCT
       S  S  S  I  K  N  R  *  *
1700 TTGAGTATGTGTGGCTAATGGTGCGTGCTCAATAAACACTTGCTG
```

Fig. 2A

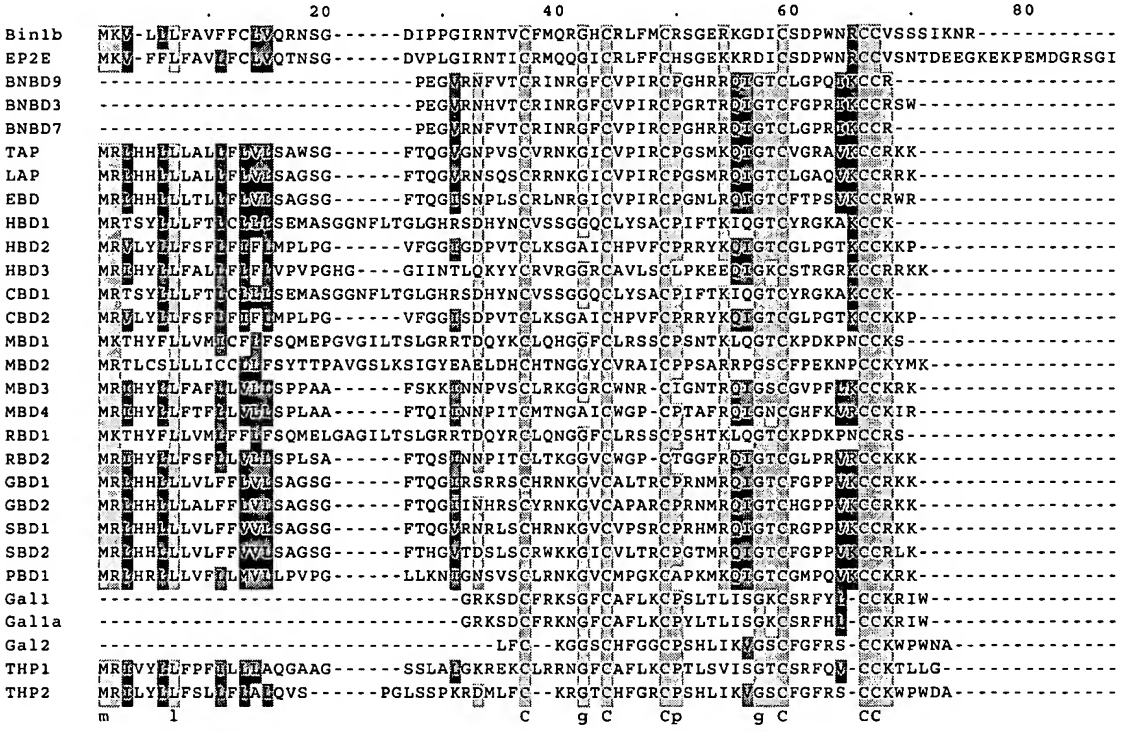

Bin1b
EP2E
BNBD9
BNBD3
BNBD7
TAP
LAP
EBD
HBD1
HBD2
HBD3
CBD1
CBD2
MBD1
MBD2
MBD3
MBD4
RBD1
RBD2
GBD1
GBD2
SBD1
SBD2
PBD1
Gal1
Gal1a
Gal2
THP1
THP2

Fig. 2B

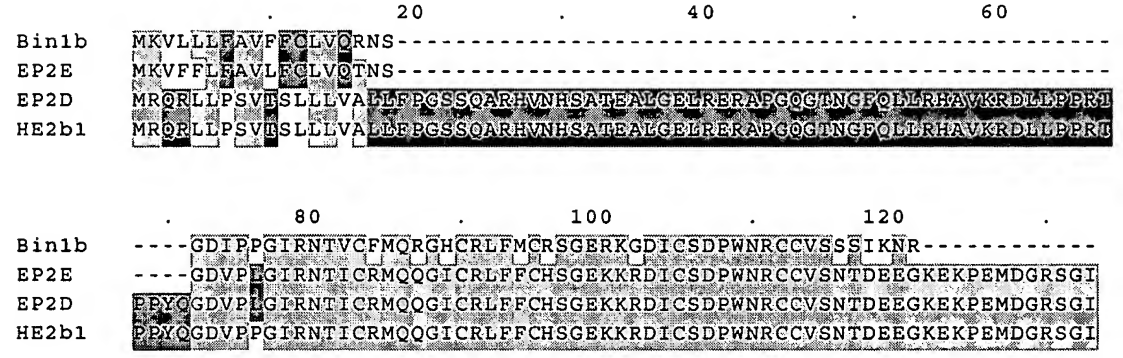
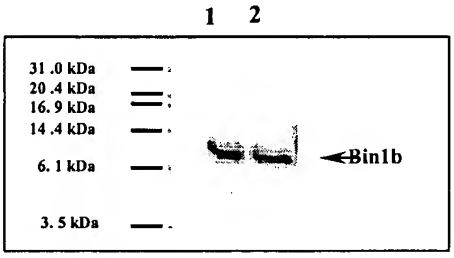
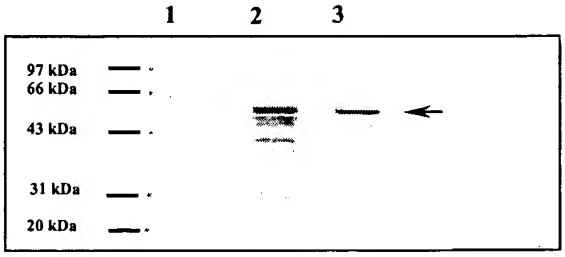

Bin1b
EP2E
EP2D
HE2b1

Fig. 2C

A.



B.



C.

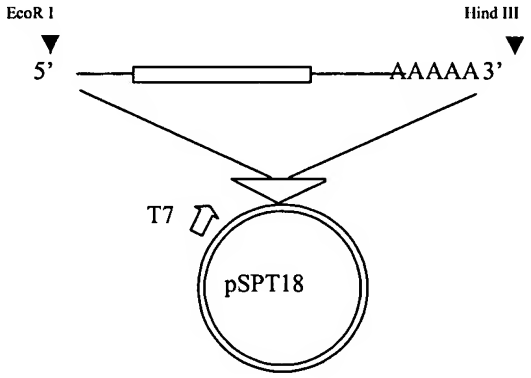


Fig. 3

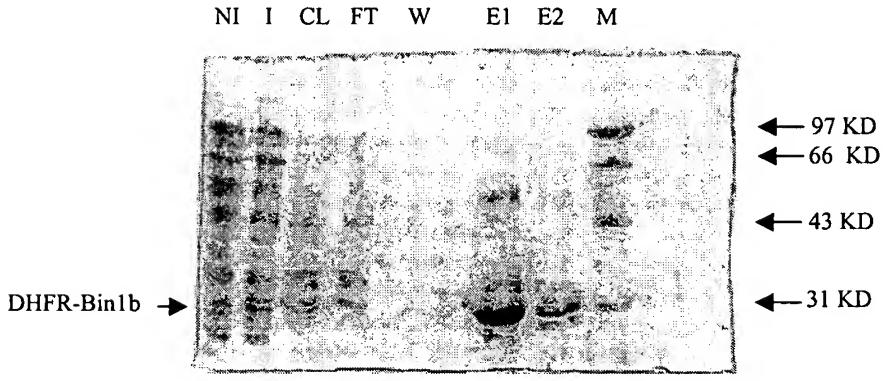


Fig. 4

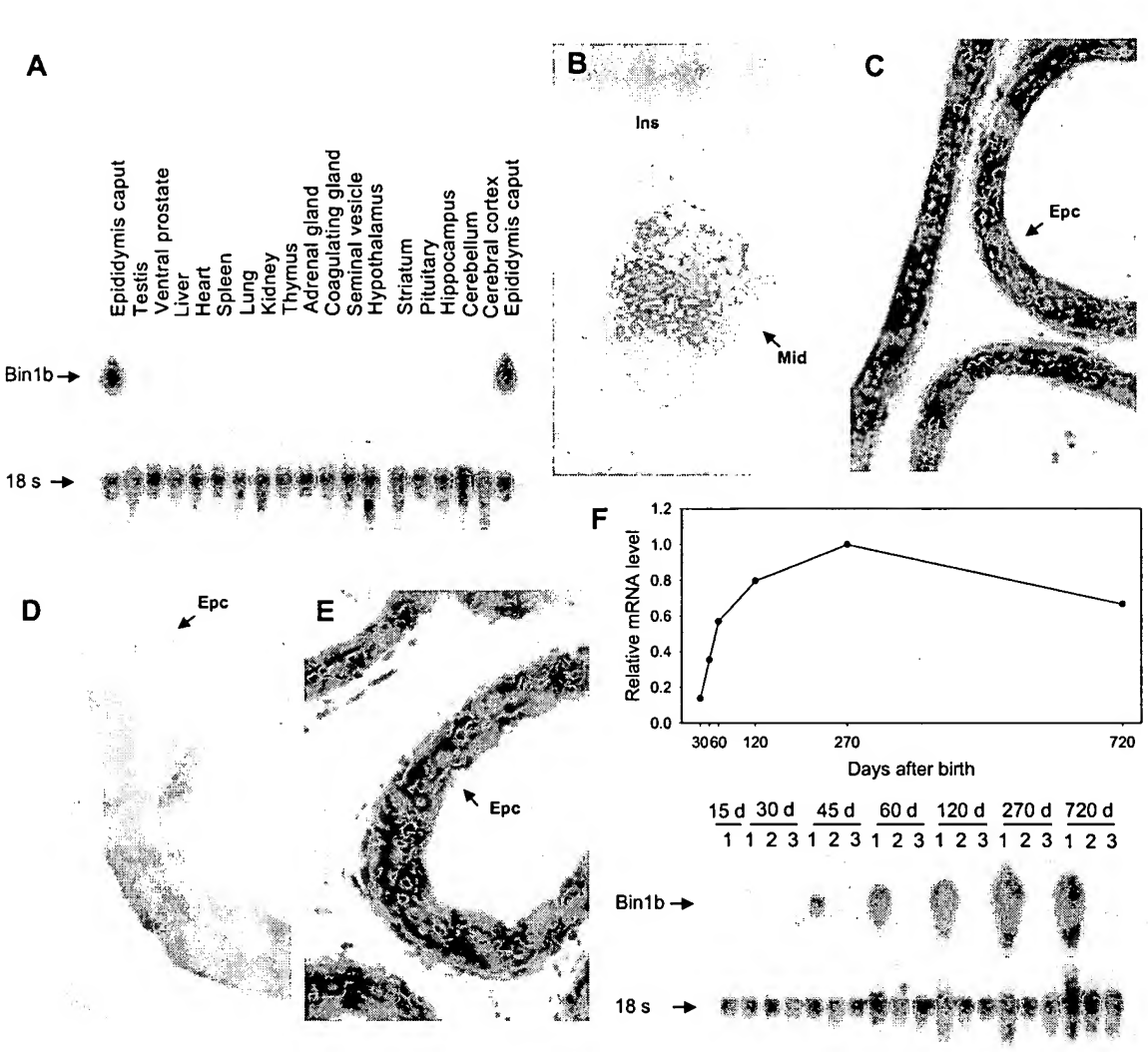
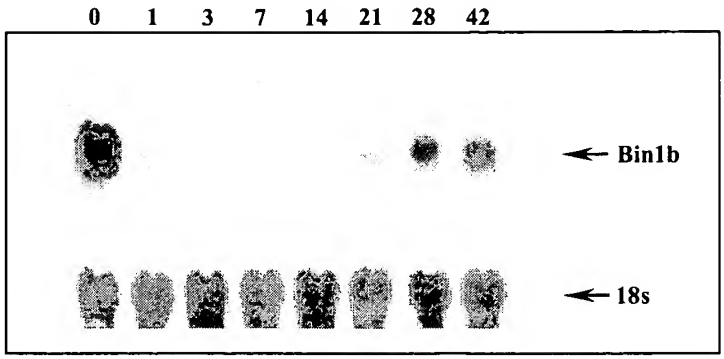

A
Epididymis caput
Testis
Ventral prostate
Liver
Heart
Spleen
Lung
Kidney
Thymus
Adrenal gland
Coagulating gland
Seminal vesicle
Hypothalamus
Striatum
Pituitary
Hippocampus
Cerebellum
Cerebral cortex
Epididymis caput
Bin1b
18 s
B
Ins
Mid
C
Epc
D
Epc
E
Epc
F
Relative mRNA level
1.2
1.0
0.8
0.6
0.4
0.2
0.0
30 60 120 270 720
Days after birth
15 d 30 d 45 d 60 d 120 d 270 d 720 d
1 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3
Bin1b
18 s

Fig. 5

A.


0 1 3 7 14 21 28 42
Bin1b
18s

B.

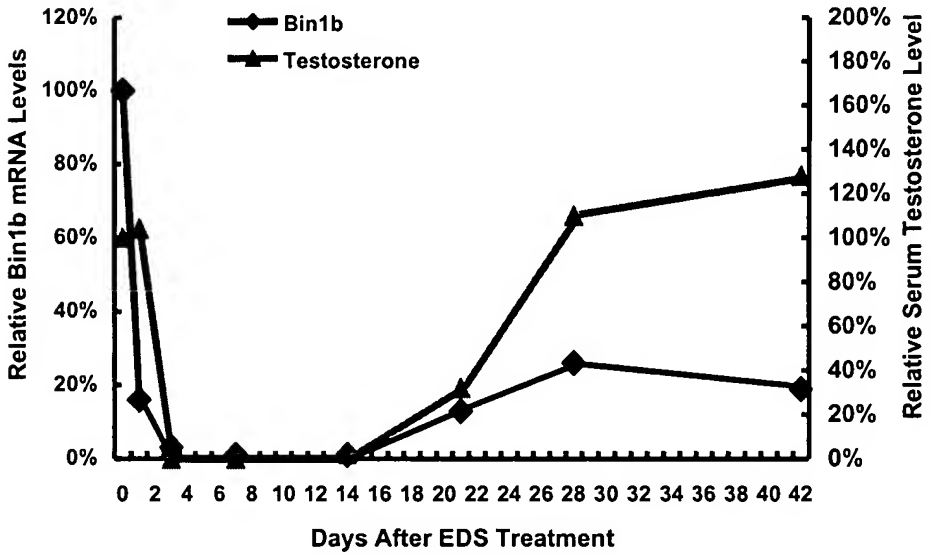

Bin1b
Testosterone
Relative Bin1b mRNA Levels
Relative Serum Testosterone Level
120%
100%
80%
60%
40%
20%
0%
200%
180%
160%
140%
120%
100%
80%
60%
40%
20%
0%
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42
Days After EDS Treatment

Fig. 6

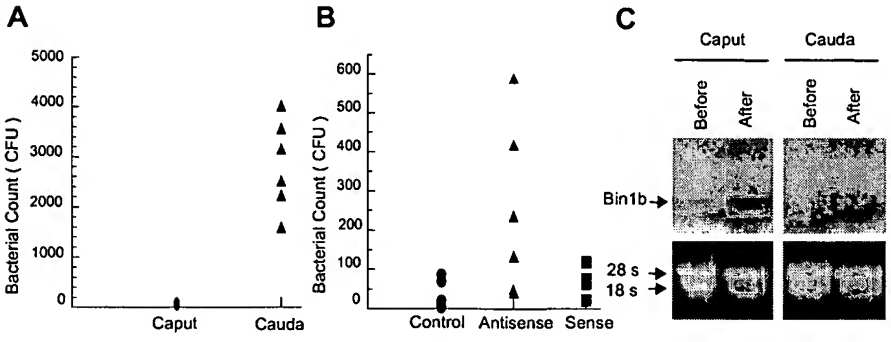

A
B
C
Bacterial Count (CFU)
5000
4000
3000
2000
1000
0
Caput
Cauda
600
500
400
300
200
100
0
Control
Antisense
Sense
Caput
Cauda
Before
After
Before
After
Bin1b
28 s
18 s

Fig. 7